DOCUMENT RESUME

ED 122 346 95 EA 008 029

TITLE	Guidelines on Pilot and Field Testing of Curriculum Products.
INSTITUTION	Contract Research Corp., Belmont, Mass.
SPONS AGENCY	Bureau of Occupational and Adult Education (DHEW/OE), Washington, D.C.
PUB DATE	75
CONTRACT	300-75-0135
NOTE	77p.
EDRS PRICE	MF-$0.83 HC-$4.67 Plus Postage
DESCRIPTORS	Curriculum Development; Elementary Secondary Education; Evaluation Criteria; *Experimental Curriculum; *Field Studies; *Guidelines; *Pilot Projects; Records (Forms); *Research Methodology

ABSTRACT

This booklet is intended to aid curriculum developers in designing and conducting pilot or field tests that yield useful information and sound evidence about the effectiveness of new curricula. Although it was prepared specifically for use by projects funded by the Curriculum Development Branch of the federal Bureau of Occupational and Adult Education, most of its content is equally applicable to field tests of other curriculum development projects as well. This publication was not developed as a handbook on evaluation, and it does not offer a precise formula for field-test design. Rather, it delineates basic principles and procedures for field testing curriculum products, presents alternative field-testing strategies, and discusses practical considerations and strategies for working with school personnel during a field test. Included in the discussion are potentially usable samples of planning and monitoring tools for pilot and field testing. (Author/JG)

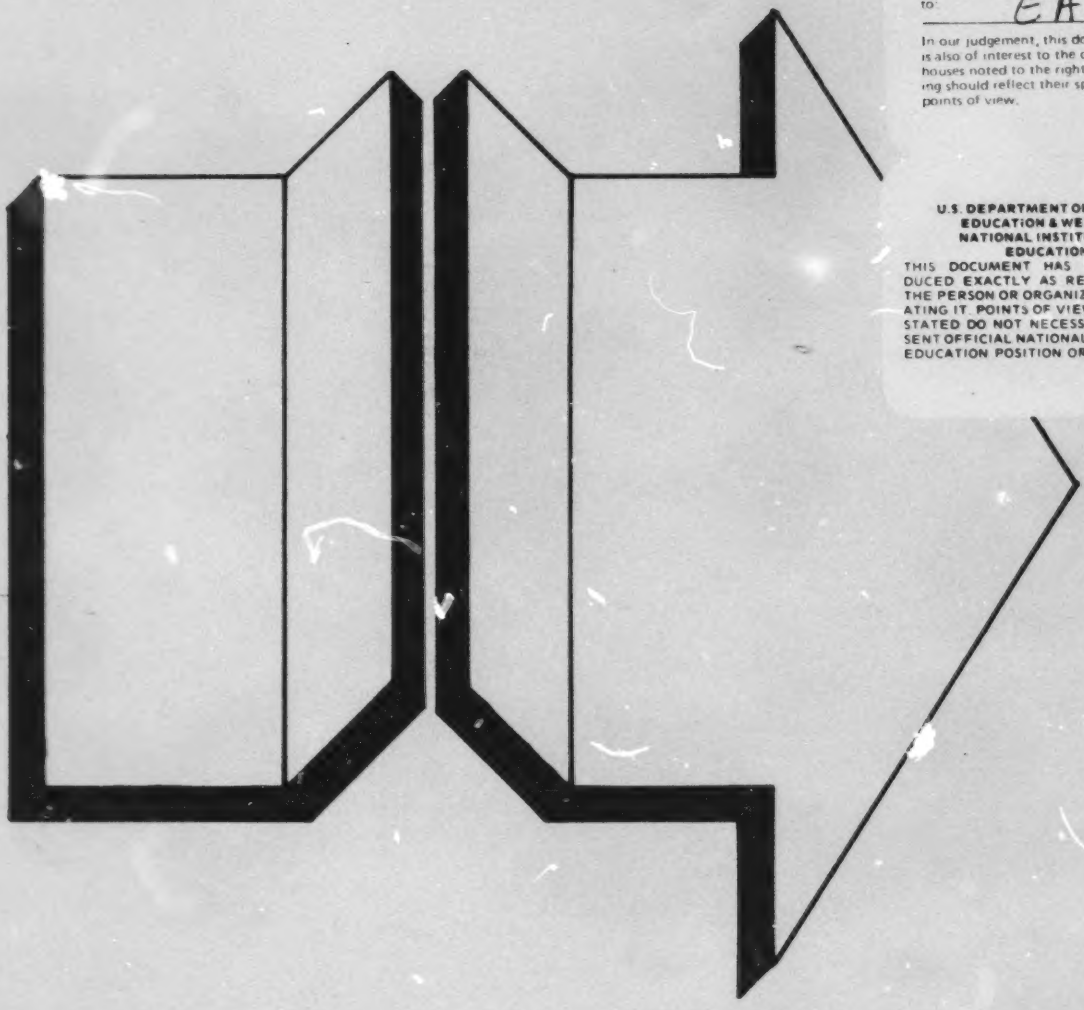
****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED122346

Guidelines on Pilot and Field Testing of Curriculum Products

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: EA CE
In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY



By
Contract Research Corporation
25 Flanders Road
Belmont, Massachusetts 02178

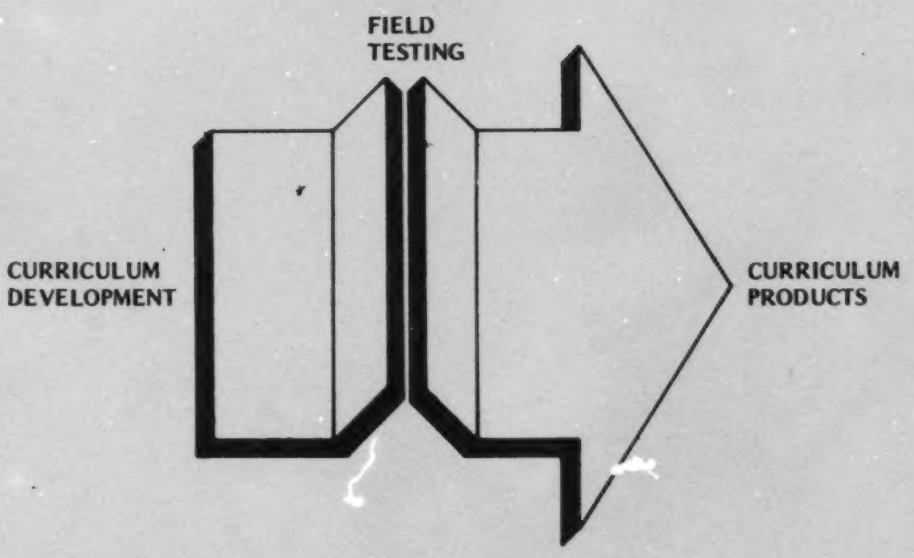
For
Bureau of Occupational and Adult Education
U.S. Office of Education
Department of Health, Education, and Welfare

EA 008 029

DISCRIMINATION PROHIBITED -- Title VI of the Civil Rights Acts of 1964 states: "No person in the United States, shall, on the ground of race, color, or national origin, be excluded from participation in, be denied the benefit of or be subjected to discrimination under any program or activity receiving federal financial assistance." Therefore, the Vocational Education program, like all other programs or activities receiving financial assistance from the Department of Health, Education and Welfare, must be operated in compliance with this law.

The project presented or reported herein was performed pursuant to a contract from the U. S. Office of Education, Department of Health, Education and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the U. S. Office of Education, and no official endorsement by the U. S. Office of Education should be inferred.

Guidelines on Pilot and Field Testing of Curriculum Products



By
Contract Research Corporation
25 Flanders Road
Belmont, Massachusetts 02178

For
Bureau of Occupational and Adult Education
U.S. Office of Education
Department of Health, Education , and Welfare

Developed and disseminated pursuant to Contract No. 300-75-0135 with the
U. S. Office of Education under Part I - Curriculum
Development in Vocational and Technical Education,
Vocational Education Amendments of 1968, Public Law 90-576

by

Contract Research Corporation
Belmont, Massachusetts
1975

PREFACE

The Guidelines were developed under U.S.O.E. Contract #300-75-0135 for the Curriculum Development Branch of the Bureau of Occupational and Adult Education. They are intended for use by projects funded by the Branch under Part I of the Vocational Education Amendments of 1968. These projects generally specify a pilot or field test of the curriculum materials.

The Guidelines are equally applicable to field test efforts undertaken in curriculum development projects supported by State or other funds.

The process of pilot and field testing is immensely challenging, involving all the technical issues of evaluation design, the complexities of relationships between developers and school personnel, and the divergent concerns of evaluation and dissemination.

To establish standards and provide a common basis of understanding among vocational education curriculum project directors, government project officers, and participating school personnel, basic principles are presented. Since product and project characteristics vary in many important ways, the Guidelines do not prescribe formulae for pilot or field-test design but suggest criteria for alternative decisions which must be made in each specific case.

The substance of the Guidelines was presented at a two and one-half day training conference held in Washington, D.C. in April, 1975. Forty Project Directors and staff representing 29 Part I projects and 15 government project officials attended the Conference. Participants represented a range of projects spanning grades Kindergarten through Adult, different product types and subject areas, and various phases of product development and field testing. The essence of the Guidelines was substantially confirmed by participants.

A large number of people contributed to the development of this document and its related training materials. These include Martin Gordon, Robert Tobin, Frederick Levy, Dr. Herbert Weisberg, Susan Bloom, Paulette Meleen and Benaree Wiley. Dr. Lou Dovner, Assistant Superintendent of the Rockland, Massachusetts Public School System, Robert Levine, Principal of Rockland Junior High School, and Cheryl Tilley, veteran field-test teacher in the Richmond, Virginia Public School System contributed their front-line perspectives to the text. Kelly Gerry, Josie Bauer, Susan Fielding, Ernest DeLuca and Barbara Boris were largely responsible for its production.

Donald Chill and Jonathan Chang, of the U.S.O.E. Forms Clearance Office, contributed information on O.M.B. clearance procedures. Dr. John Evans, Director of the Office of Planning, Budgeting and Evaluation, U.S.O.E., described the purposes and activities of the Dissemination Review Panel which he chairs. Dr. Howard Hjelm, Director of Research, Division of Vocational-Technical Education, provided insights into issues of dissemination. Mary V. Marks, Chief, Curriculum Development Branch, lent continuing and helpful support.

Alice K. Gordon
Project Director
September, 1975.

GUIDELINES ON PILOT AND FIELD-TESTING OF CURRICULUM PRODUCTS:

Principles, Alternative Strategies, Decision Criteria, Policies and Procedures

TABLE OF CONTENTS

Chapter		Page
Introduction		1
1.0	Purposes of Pilot and Field Testing	3
1.1	Goals and Objectives of Testing Curriculum Products	3
1.2	The User Framework: Results of Testing	5
1.3	The Field-Test Process: Achieving Consensus	6
1.4	Field-Test Findings	7
	Summary	9
2.0	Overview: Scheduling the Pilot or Field Test	10
2.1	Overview	10
2.2	Planning Calendar	10
	Summary	13
3.0	Design and Planning of a Pilot or Field Test	14
3.1	Specifying Research Goals	14
3.1.1	Hierarchy of Goals	15
3.1.2	Pilot and Field-Test Goals	16
3.1.3	User Standards of Evidence	17
3.1.4	The Dissemination Review Panel	18
3.1.5	Resource Constraints	19
3.1.6	Pilot and Field-Test Objectives: Formulating the Question	20
3.2	Developing the Design	25
3.2.1	Working with Givens	25
3.2.2	Basic Elements of the Design	25
3.2.2.1	Curriculum	26
3.2.2.2	Students	28
3.2.2.3	Teachers	30
3.2.3	Establishing Standards of Comparison	30
3.2.3.1	Randomized Control Group	31
3.2.3.2	Non-Random Control Group	31
3.2.3.3	No Control Group	32
3.3	Developing a Data Collection Plan	33
3.3.1	Formulating Hypotheses	34
3.3.2	Defining Information Needs	34
3.3.3	Selecting a Data Collection Technique	35

Page

3.3.3.1 Tests and Measurement Instruments 35
3.3.3.2 Observation 36
3.3.3.3. Interview 36
3.3.3.4 Questionnaires 36
3.3.3.5 Existing Records 36

3.3.4 Identifying Conditions and Constraints 36

3.4 Securing OMB Clearance 37

3.4.1 Documents Requiring Approval 37
3.4.2 Material to Be Submitted for Clearance. 38
3.4.3 The Supporting Statement 38
3.4.4 Time Typically Required for Clearance 39

3.5 Formulation of Criteria and Policies 39

3.5.1 Criteria for Site Selection 39
3.5.2 Field Test Policies 40
Summary 41

4.0 Site Seeking, Finding, Holding 44

4.1 Overview of the Process 44
4.2 Identifying Potential Candidates 44
4.3 Contacting Potential Sites 45
4.4 Reaching Consensus 49
Summary 50

5.0 Implementation of the Pilot or Field Test:
Interaction Between the Developer and the Site 52

5.1 Pre-Implementation Activities: Teacher Preparation 52
5.2 On-Site Monitoring 56
5.3 Post-Implementation: Closing the Loop 58
Summary 59

6.0 Data Analysis and Interpretation 60

6.1 Comparing Experimental and Control Outcomes 60
6.2 Item Analysis 63
6.3 A Practical Consideration 63
Summary 64

7.0 Use of Results 65

7.1 Curriculum Revisions 65
7.2 Dissemination Decisions and Procedures 66
Summary 69

INTRODUCTION

Purposes of the Guidelines

The Guidelines were written to:

a) assist curriculum developers to design and conduct pilot or field tests which yield useful information and sound evidence about the workability and effectiveness of a new curriculum.

b) acquaint all actors, but primarily the initiating developer, with the context of field tests and the needs and responsibilities of all parties to it.

c) promote the adoption by sponsors* of policies and conditions which support sound field test plans.

The Guidelines therefore:

a) delineate basic principles and procedures which meet standards of quality research for testing curriculum products in the "setting of use."

b) present alternative strategies for pilot and field testing of curricula in terms of decision criteria for developing a specific plan suitable to the characteristics of the curriculum to be tested and the resources available.

c) describe practical considerations and strategies for approaching, securing and working with school personnel upon which the success of a pilot or field test depends.

d) include potentially usable samples of planning or monitoring tools for pilot and field testing.

What the Guidelines Are Not

The Guidelines were not developed as a handbook on evaluation. Technical treatises of this type abound and the problem of pilot and field testing extends beyond issues of research methodology.

The Guidelines do not offer and do explicity reject a single recipe or formula for field test design. Curriculum projects, particularly in career and vocational education, vary along a number of important dimensions, including:

- the objectives of the field test
- the type of product or nature of the program to be tested
- the intended users of the curriculum
- the resources for field testing
- the evaluation design constraints
- dissemination criteria
- other factors, including access to data and conditions imposed by the sponsoring agency

* Sponsor refers to Federal or state funding agencies.

These differences are not mere variations on a theme but tend to be decisive for field-test designs. Thus, basic principles of sound research and procedures must be applied to accommodate the specific characteristics of a given curriculum development project. No one approach will suit all situations. A thoughtful plan must be tailored to each.

The Guidelines assume two basic commitments:

1. A field-test design must strive to meet standards of sound research.

2. A field-test plan must seek attainable objectives, realistic for the resources available and other conditions.

The two commitments often conflict. Standards of sound research may not be attainable within a given project; attainable plans may not satisfy standards of evidence. Each field-test effort must therefore reflect a negotiated solution between these two goals.

These Guidelines address both desirable standards for pilot and field testing and considered adjustments or compromises from those standards. Increased understanding by developers, sponsors and practitioners of the requirements for establishing evidence of effectiveness should help to promote improved conditions for field testing so that imcompatibility between desirable research standards and the realities of test conditions will be reduced.

SUMMARY

1. The purpose of these Guidelines on pilot and field testing is to assist curriculum developers to accommodate principles of sound research on the one hand, the realities of resource constraints, practitioner needs, dissemination/adoption standards of different users, on the other.

2. A field-test design must strive to meet standards of sound research.

3. A field-test plan must seek attainable objectives within the resource constraints.

4. The Guidelines on pilot and field testing provide decision criteria and considerations rather than prescriptive statements or recipes for evaluation.

Chapter 1. Purposes of Pilot and Field-Testing

No matter how attractive or technically accurate, new vocational-technical curricula cannot be disseminated or adopted confidently without information about their usability and effectiveness with the target population. The curriculum developer cannot tell from behind a desk whether a curriculum will "work." The potential user is justifiably reluctant to adopt a new curriculum solely on the basis of examination or on the testimonial of the developer. For both the producer and consumer of new curricula, pilot and/or field testing of curriculum materials serves an essential function. The developer can improve curriculum materials based on experience gained from actual trial in the classroom, or other intended "setting of use." The consumer can make decisions to adopt based on empirical evidence of effectiveness gained in this way. Pilot and/or field testing therefore represent an essential phase of curriculum development projects. By encouraging use of demonstrably effective curricula, improvement of instruction through curriculum development may be achieved.

1.1 Goals and Objectives of Testing Curriculum Products

Pilot and field testing are therefore instrumental to dissemination of new curricula. Two basic goals of curriculum testing are:

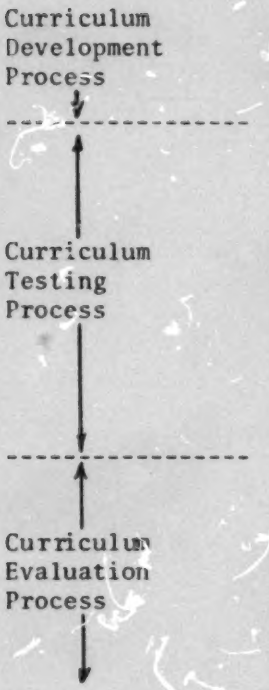
1. to improve curriculum products
2. to establish evidence of effectiveness

Involvement of curriculum staff in product development activities tends to focus attention on development goals. The satisfaction of meeting substantive, aesthetic, and other product criteria may obscure the dissemination purposes for which curriculum development is initiated by Federal, state and other sponsors.

At least theoretically, evidence that a curriculum achieves its objectives is an essential prerequisite to dissemination. Neither the developer nor the adopter has sufficient rationale for promoting or adopting new curricula without such evidence. In reality, other motives may influence user decisions but these then operate at the level of a marketplace of products, offering choice of styles and types, outside the arena of research and development activities.

Pilot and field testing may be understood as mid-range research efforts on a scale of types of testing, as shown in Table 1-A following:

Table 1-A. Scale of Testing of Curriculum Products



Product Verification	Judgment of professionals and of employers/employees that a curriculum is accurate, comprehensive, relevant, clear, etc.
Developmental Testing	Trial use of part or all of a curriculum by the development staff on the target learner population or a substitute.
Pilot Testing	Trial use of part or all of a curriculum in the "setting of use" generally in one site and/or for mostly formative objectives (workability).
Field Testing	Trial use of the curriculum in the intended "setting(s) of use," in multiple sites, for formative and summative (workability and effectiveness) objectives.
Evaluation	Use of the curriculum in many "settings of use" primarily to measure effectiveness.
Evaluation Follow-Up	Longitudinal or follow-up study of impact of curriculum on student performance or decisions over time in other settings (e.g., employment). In effect, measures relevance or efficacy of curriculum objectives for its ultimate goals.

The definitions and distinctions above were developed to characterize typical patterns of curriculum assessment found among a sample of 30 different projects funded under Part I of the Vocational Education Amendments of 1968. They do not necessarily represent widely-used terminology, which was either unavailable or insufficiently detailed for the vocational curriculum projects addressed in these Guidelines.

The terms and definitions are operative in the Guidelines. Several important features of them should be noted.

1. Product Verification is a part of the development process and does not constitute a test of curricula in the "setting of use."

2. Each category may embrace more than one type of approach to curriculum testing but is disqualified in some aspect from a more rigorous category. This is particularly true of the term pilot testing which may be identical to field testing in all respects except limited to one site.

3. Distinctions of purpose (e.g., curriculum testing vs. curriculum evaluation) represent general rather than sharp distinctions. Field testing, for example, may address curriculum evaluation as well as less rigorous curriculum testing. Development testing may be considered part of the development process but shares curriculum testing objectives to some degree.

4. Pilot and Field Testing are the focus of these Guidelines. Curriculum evaluation is desirable but difficult to achieve under typical curriculum development project conditions. Longitudinal evaluation or follow-up studies are generally beyond the scope of Part I projects.

Under conceptually ideal conditions, curriculum product development and evaluation would proceed through each type of testing as a distinct time frame. For example, materials would be tested and revised at each step before being tested at a bigger scale. Project staff would test a curriculum on themselves before attempting to pilot it in a classroom. This system of feedback loops is generally too costly in time and money for implementation and may delay dissemination to users to the point of jeopardizing the relevance of vocational curricula. This is the central dilemma of the relationship between evaluation and dissemination.

The realities of field-test conditions limit these Guidelines primarily to pilot and field testing as defined in the scheme above in Table 1-A.

1.2 The User Framework: Results of Field Testing

The key use of evaluation, of which pilot and field testing is a subset or special type, is to furnish information for decision-making. The specific choices and criteria, however, vary with different user groups.

The types of decision faced by different parties interested in vocational education curricula and their decision criteria or standards of evidence are briefly illustrated in Table 1-B below.

Table 1-B.

User	Decisions	Standards of Evidence
Sponsors	Continued funding Dissemination	Rigorous evidence of curriculum effectiveness on adequate sample populations
Developers	Curriculum revisions	Judgmental evidence from small numbers of teachers and/or students serve as valid bases for revisions.
Administrators	Adoption and utilization	Evidence of effectiveness of curriculum for specific target populations; of usability by teachers. Teacher judgment may be as important as information on student achievement.
Teachers	Utilization	Evidence of usability; motivational qualities of curriculum; prerequisite knowledge and resources for teaching; target populations and effectiveness.
Students	Enrollment may or may not be a choice	Expected outcomes and prerequisites; relevance to post-educational goals.
Parents	Adoption in schools Student enrollment in program	Expected outcomes and prerequisites; relevance to post-educational goals
Employers	Employment during or post-program of enrollees or graduates	Evidence of effectiveness; relevance of curriculum objectives to performance outside the program.

The different concerns represented require either different types of field tests or different emphases. Some information needs cannot be easily met. For example, employer interest in on-the-job performance of students trained in a particular program would require a follow-up study beyond the scope of field testing. A suitable substitute might be formed by a combination of employer verification of curriculum content and objectives and evidence of student achievement of objectives even for small samples (e.g., one class). Dissemination criteria employed by sponsor agencies* require rigorous evaluation standards, including multiple sites and adequate control groups.

Defining the User Framework

- The field tester must determine in advance the appropriate decision-makers, decisions, decision criteria, and the associated information needs which will be addressed in the field test.
- Divergent concerns of user groups may not be reconcilable in a given field-test design. Definition of users is essential.

1.3 The Field-Test Process: Achieving Consensus

Analogous to divergent interests of users of the end product of a field test are the different concerns of participants in it. The field-test process must be understood in its operational context. The curriculum developer is not a lone technician seeking a solution to a technical evaluation problem. Rather, he/she depends upon a variety of other actors who participate in an evaluation plan and have a stake in it. For each of these actors, the objectives and expectations of a plan or field test may differ. Each expects to gain something from participation in the field test, but few have a research perspective or objective. To the extent that these different agenda do not correspond to the evaluation requirements for establishing evidence of effectiveness, the field test process stands in serious jeopardy.

- Curriculum Developers generally seek to determine a) the usability and b) the effectiveness of the curriculum.
- Teachers participating in a field test generally have an action rather than a study agenda. Reasons for participating include: a) desire to install a new subject area or improve an established one b) inclination or need to cooperate with a requirement imposed by the school administration with little personal commitment to it c) desire to try something new. Conditions of participating may include: a) interest in only portions of the curriculum b) intentions to participate as long as the curriculum seems to "work." Divergence with field-test objectives occurs particularly when field testers rely on teachers as key sources of data for which the teacher has no intrinsic use.

* U. S. Office of Education dissemination criteria, as reflected in the decisions of the Dissemination Review Panel, are discussed in Chapter 3, page 18.

- Administrators often seek a) visibility for their schools b) staff development c) early fulfillment of new curriculum plans in their systems.

- Students rarely have the opportunity to express a choice or objective for participating in a field test unless the field test is conducted in an elective course with voluntary enrollment. Students do, however, have a stake in a field-test effort, especially to the extent that a) the materials are ineffective in some respect but the evaluation design requires that they proceed b) the test uses classroom time which could be spent on subject areas more important to them c) additional time is required of the students for the sake of the field test. Curriculum developers therefore have the responsibility to modify materials and/or the evaluation plan in light of student needs.

- Sponsors generally seek "hard" evidence that a curriculum is effective in order to justify funding decisions or meet dissemination criteria.

In short, the individuals responsible for implementing the field test have the greatest opportunity and perhaps the greatest need to depart from an evaluation plan. This problem, not unique to educational evaluation, emphasizes that a technically sound evaluation plan, while very important, will not in itself assure a successful pilot or field test. The field tester must bring into congruence the divergent expectations of the actors involved. This requires making certain that all parties to the field test understand its purposes, processes and demands and consent to their roles in meeting them.

1.4 Field-Test Findings

The benefit of hindsight must be sought at the outset of the field test. In planning a field-test strategy, it is essential to project the desired outcome of the research. Assume the position temporarily and hypothetically of a developer reporting out on a completed field test. This necessary process of making explicit the end products of the field test relates directly to the user decision needs which the field tester hopes to serve.

For the curriculum developer in vocational education, the process is a familiar one. It is analogous to defining the competencies that a student must have for employment prior to specifying the learning objectives, content or instructional strategies required to develop such competencies.

For pilot and field testing, specifying the program goals which the field test seeks to measure must occur prior to developing a field-test plan. One approach to operationalize this process is to construct at the outset a statement of research results or "abstracts" comparable to those which the developer hopes to make at the conclusion of the test. This does not, of course, entail prejudging the success of the curriculum but rather of clearly specifying the type of findings sought. Typical types of field-test findings are illustrated in the four abstracts which follow.

Abstract #1: Usability

The complete curriculum was used, without developer assistance, by ten different classroom teachers. No serious problems in its implementation were reported. Teachers reported that the materials were clearly worded and readable by students; instructional strategies outside the classroom were feasible; and that student motivation was positive as indicated by participation in community exercises, active class discussions, and timely completion of homework assignments.

Abstract #2: Evidence of Effectiveness

The curriculum successfully trains students for the 5 specified occupations in this cluster. Of the 50 students enrolled, 80% achieved the performance objectives established in the curriculum as demonstrated on written and observational tests provided in the curriculum.

Abstract #3: Evidence of Effectiveness

The curriculum successfully increases students' understanding of and interest in career options in this field. 75% of the 300 students enrolled in the curriculum were able to describe employers, working conditions, occupations and job tasks, responsibility and rewards for 30 occupations in this cluster. Only 5% of the students could describe 10 occupations in these terms prior to taking the course.

Of the 300 students, 90% participated in a hands-on work experience program for a minimum of 2 weeks.

Of the 300 students, 45% expressed interest in further study and possible careers in this field as compared to 2% who did so at the outset of the program.

Students were randomly selected for the program.

Abstract #4: Relevance

The curriculum successfully prepares students for careers in this field. 95% of the 50 students enrolled secured jobs in the field for which they were trained. Employers verified competent on-the-job performance by all of these students. One-year later, 92% of the students were in the same job or had been promoted by the same employer.

Students reported that their training had prepared them well for their jobs and indicated job satisfaction. This compares favorably with student employment patterns, employee and student satisfaction prior to the introduction of this curriculum.

The type of statement the developer wishes to make at the conclusion of the test (methodological considerations momentarily aside) must direct the choice of field test.

SUMMARY

1. Pilot and field testing refer to trial uses of the curriculum in the intended "setting of use."

2. Product verification through professional judgment is part of the curriculum development process, not the testing process.

3. User decision criteria influence the choice of field-test objectives.

4. Participants in a field test have different agenda; these agenda must be brought into congruence to foster a successful pilot or field test.

5. The type of conclusions sought about a curriculum should be identified as the first step in planning the pilot or field-test design.

6. Pilot and field testing are instrumental to dissemination of new curricula.

Chapter 2. Overview: Scheduling the Pilot or Field Test

2.1 Overview

Pilot and field testing involve a set of decisions and operations which proceed in five phases, as follows:

1. Design and Planning of a Pilot or Field Test
2. Site Seeking, Finding, Holding
3. Implementation of Pilot and Field Test: Interaction between Developer and Sites
4. Data Analysis
5. Use of Results

These phases, though they partially overlap in time, represent the sequence in which planning and implementation of a pilot or field test occurs. Subsequent chapters of the Guidelines correspond to each of the five phases in turn.

This organization of the Guidelines differs from the usual treatment of educational evaluation divided into three segments: research design, data collection, and data analysis. The difference reflects a special feature of field-test efforts which differentiate them from more typical educational evaluations.

Pilot and field-test efforts cause the use of a curriculum for the sole purpose of assessing it. Responsibility for promoting initial implementation of a curriculum lies with the developer/field-tester staff. In the more usual situation, an evaluation is undertaken external to the program and no responsibility for initiating the program is borne by the evaluation staff. Curriculum testing thus is best discussed in sequential terms to examine the important relationships between the study and the action features.

2.2 Planning Calendar

Most projects are understandably reluctant to plan a pilot or field test in any detail until curriculum development is complete. The schedule of Part I projects and the typical school calendar, however, make this sequencing impractical.

Early planning is essential to assure effective pilot and field-test efforts. The developer must establish a schedule which allows for timely completion of all curriculum test tasks.

Several key events determine critical milestones in the planning process:

- the initiation date of the pilot or field test itself, usually coinciding with the beginning of a school semester

- the date by which approval for data collection activities from the Office of Management and Budget (OMB) must be secured

- the date by which submission of OMB clearance package must therefore be accomplished (generally 3 months prior to need for approval)

Equally important but somewhat more flexible dates will be determined by:

- the time by which teachers must receive curriculum materials and/or training in the use of the curriculum

- the time by which teachers and therefore sites must be secured

Where recruitment of instructors or considerable investment by schools is required, especially at the postsecondary level, field-test planning must begin simultaneous with curriculum development.

A pilot and field-test planning schedule which accommodates these key milestones will generally overlap with curriculum development tasks. The curriculum project itself, however, may be enhanced if questions pertinent to the pilot or field test, such as measures of student performance or classroom hours required for the curriculum, are addressed early in the development phase. Early planning also enables the developer to make rational allocations of resources between developmental and testing activities.

The developer needs to construct a schedule tailored to the nature of the curriculum, the project duration, and the requirements of the field test. A typical pattern for field testing of Federal vocational curriculum projects, assuming funding 1 July and completion 30 June two years later, follows.

Table 2-A.
Suggested Timetable for Curriculum Field-Test Planning
For Two-Year Projects*

	Task	To Be Completed By Beginning of Month
1.	Develop Field-Test Plan Establish Criteria for Test Schools	7
2.	Begin Contact with Candidate Schools, Secure Data Necessary for Selection Decisions	8
3.	Select Schools and Teachers	9
4.	Complete Monitoring Instruments and Data Collection Forms	11
5.	Submit OMB Clearance Package to Project Officer	11
6.	Deliver (at least a Portion of) Draft Curriculum to Test Teachers	11
7.	Train Teachers	13
8.	OMB Approval Secured	14
9.	Field-Test Initiation: Pre-Test Administration	15
10.	Implementation Phase: Monitoring, Technical Assistance, Data Collection	16-18
11.	Post-Implementation Phase: Complete Debriefing, All Data Collection	19
12.	Data Analysis	20
13.	Curriculum Revisions (ongoing)	22
14.	Sponsor Review (ongoing)	23

*Many projects may accomplish certain tasks prior to month indicated in timetable. Times indicated are latest dates for task completion to ensure proper field testing.

SUMMARY

1. The pilot and field-test process proceeds through five phases:

   a) Design and Planning
   b) Site Seeking, Finding, Holding
   c) Implementation
   d) Data Analysis
   e) Use of Results

2. Pilot and field testing differ from conventional educational evaluation in that implementation of the curriculum is initiated for the purpose of assessing it.

3. Pilot and field testing require advance planning.

4. Specific schedule milestones are determined by lead-time for CMB clearance; teacher recruitment/training requirements, etc., calculated back from the initiation date of the field test.

5. Such planning needs to be accomplished in parallel with curriculum development efforts.

6. For certain types of curricula, particularly postsecondary, field-test planning must begin with the initiation of curriculum development.

Chapter 3. Design and Planning of a Pilot or Field Test

The design and planning of a field test of curriculum products serves as the basic foundation of the entire field-test phase of curriculum development. The more thoughtful, thorough and careful the design, the more successful the pilot or field test will be. Design and planning involves 6 basic elements:

1. specifying the research goals
2. developing the pilot or field-test design
3. developing a data collection plan
4. securing OMB clearance
5. formulating criteria and policies

3.1 Specifying Research Goals

The pivotal decision the developer must make is what he/she wishes to learn from a test of the materials. The general implicit purposes of curriculum testing do not constitute sufficient direction for a pilot or field-test effort. Nor do the assumed assessment procedures (e.g., multiple sites; pre- and post-testing), often specified contractually,* constitute a pilot or field-test plan. The development of a productive research design derives from the goals specified for the inquiry, and the developer must be able to articulate these goals to all participants for the purpose of obtaining consensus.

Ideally, the goals of the curriculum test should dictate the choice of pilot or field-test plan, not the other way around. In practice, resource constraints of time, funds or product type limit the goals which may be attained in a pilot or field test. For example, where the curriculum consists of six 1-hour modules for infusion over 3 years of classes, the usefulness of measuring student learning may be questioned.

In designing a pilot or field-test effort, a rational process for prioritizing research goals must be employed in order to produce a useful, achievable design.

* Specifications of number and type of sites in a Federal Request for Proposal for curriculum development projects might best be interpreted as an indicator of the feasible limits of a design, not its specific form. (The sponsor's relative interest in development versus testing of the curriculum might also be inferred. Alternative allocations of project resources to development activities within a given project might be considered. The overwhelming emphasis on materials development might be reduced in favor of increased testing of fewer materials.)

3.1.1 Hierarchy of Goals

The goals of a pilot or field test must:

- correspond to the goals and objectives of the curriculum
- respond to the standards of evidence acceptable to decision-makers or users of the results
- accommodate resource constraints

A curriculum is designed to accomplish a hierarchy of goals. The bottom range of the hierarchy includes many specific, concrete, immediate objectives. The top range of the hierarchy addresses more abstract, long-range goals. The achievement of higher level goals is dependent at least in part on the achievement of more proximate ones.

A simple example of such a goal hierarchy is as follows.

TABLE 3-A

Goal Hierarchy

- Enhance quality of life
- Increase developmental potential of individual
- Provide individuals with adaptability skills
- Enable individuals to obtain satisfying employment
- Place students in postsecondary training
- Train students to perform occupational skills
- Enable students to select careers in occupational fields
- Increase student awareness of career and training opportunities
- Offer students an educational opportunity in the occupational cluster
- Implement a new offering in schools

Evaluation seeks to measure the degree to which program goals are met. Generally, educational evaluation seeks the appraisal of the outcomes of student learning. Evidence of effective student learning is usually accepted as a proxy for more ultimate program goals. For example, if students demonstrate increased awareness of career and training opportunities in a secondary vocational curriculum, it is possible to infer that students will proceed to obtain postsecondary training.

To conclude from this evidence, however, that the employment goals of the curriculum (i.e., that students get jobs) have been met, requires an inferential leap for which the evidence is weak or insufficient.

If, however, an evaluation confirmed that students who had achieved the learning objectives of the curriculum had proceeded to obtain postsecondary training, the strength of evidence of effectiveness would be increased and the inference that the curriculum attained its employment goals more justified.

The goal hierarchy serves as a means of conceptualizing a series of tests through which a curriculum or program must pass in order to demonstrate increasingly powerful evidence of effectiveness. Evidence of achievement of each goal in the hierarchy represents necessary but insufficient evidence of achievement of succeeding goals.

Pilot and field testing of vocational-technical curricula generally seek to measure achievement of mid-range goals. In this respect, they share characteristics of most educational evaluations which seek evidence of student learning as indicative of program success, one basis for funding, dissemination and adoption decisions.

3.1.2 Pilot and Field-test Goals

Pilot and field tests can be undertaken to assess two types of curriculum goals:

a) the extent to which the curriculum can be implemented

b) the extent to which students achieve the learning objectives of the curriculum

The developer can attempt to assess these goals in sequence or at once, depending upon the purpose of the pilot or field test and predictions about its probable success.

A pilot or field test may be undertaken simply for the purpose of gaining information to improve the curriculum products. This type of inquiry is generally referred to as formative evaluation. It is particularly suitable for conditions such as the following:

- a curriculum with entirely new content or unconventional instructional methods, including extensive out-of-school activities
- a curriculum which requires implementation strategies not usually employed in schools
- a curriculum which does not fit the existing division of courses in schools
- a curriculum which does not constitute a total program

Presumably, a curriculum which passed the test of implementation or usability would subsequently be evaluated in terms of student learning outcomes.

An alternative strategy is to seek information about both goals above simultaneously in one inquiry. Such an assessment undertaken in part for the purpose of assessing program effectiveness is referred to as summative evaluation. The developer may reasonably attempt to examine questions of usability and of student learning if, for example:

- portions of the curriculum have already proven usable
- its methods are well-established
- professional judgment (e.g., Advisory Panel) verifies curriculum content and methods and predicts it will succeed

3.1.3 User Standards of Evidence

User standards of evidence influence the type of inquiry a developer must undertake to satisfy user decision criteria. Standards of evidence refer to:

a) acceptable levels of inference

b) standards of rigor

Levels of inference refer to the scale of criteria necessary for decision making. Users who will not consider, for example, adopting a curriculum until its effectiveness for placing students in jobs has been verified, almost certainly cannot be satisfied by field-test conditions. On the other hand, adopters of curricula who consider evidence of usability as sufficient basis for concluding that the curriculum achieves its objectives do not require that the field tester appraise student learning at all. This minimal standard of evidence requires a gigantic inferential leap which some potential users of field-test results are legitimately unwilling to make.

The field tester, as a participant in research and development activities, has an obligation to attempt to satisfy the highest user standards of evidence possible within resource constraints.

Resource constraints may have more influence on the power of conclusions about a curriculum than on the type of conclusions themselves. For example, even in small-scale pilot tests, evidence of student learning may be obtained. The extent to which such evidence may be generalized to the total population, however, may be nil. In general, it is better to seek evidence of student learning on a small scale than to settle for verification of usability on a large scale.

Standards of rigor refer to the degree of confidence with which conclusions about a program may be asserted. The extent to which effects may be attributed to a curriculum rather than to other factors depends primarily on procedures and characteristics of the evaluation design. Defects in such procedures increase the plausibility of alternative explanations of the effects and decreases the strength of evidence for program effects.

User standards of rigor vary. The function of the decision-maker influences to some degree the standards of rigor appropriate. In research and development

activities, the sponsor is likely to uphold the highest standards of rigor because responsibilities for dissemination of a program or curriculum to encourage replication of it in other settings cannot be undertaken without convincing evidence of effectiveness.

3.1.4 The Dissemination Review Panel

At the Federal level, the U.S. Office of Education Dissemination Review Panel is a formal arbiter of such standards. To earn its approval for dissemination, a program or project must produce objective evidence that it made a difference and that such difference could be attributed to the effects of the program rather than to something else.

General criteria employed by the Panel are useful to examine. They reflect the legitimate expression of accountability required of innovators and promoters of change. The considerations below apply mostly to programs aimed at increasing student cognitive or achievement gain. Other kinds of criteria suit other types of programs.

TABLE 3-B

U.S.O.E. Dissemination Review Panel Concerns

1. Measures are taken before and after the program which demonstrate student gain or improvement.
2. The gain is large enough to be meaningful in statistical and educational terms.
3. The results of the program can be generalized. This requires adequate sample sizes.
4. The program itself is defined and was implemented in more than one setting.
5. Gains or improvements can be attributed to the project rather than to other factors, including normal maturation of the student. A measure or estimate of what would have happened to students if they had not been in the program, obtained from control groups or other methods, establishes the comparison.

The degree to which pilot and field testing of vocational-technical curricula may meet standards of rigor will vary with each situation. The conditions of pilot and field testing, however, are often inadequate for meeting such standards. These conditions include:

- limited numbers of sites, schools, classes, students;
- absence of standards or norms for the type of gains these curricula try to produce and consequent difficulty of judging significance of gain;
- brief duration of the field test.

The highest standards of rigor attainable in a given situation should be met. However, what these standards are and limits on them should be made

explicit in each case so that sponsors and developers share clear and reasonable expectations for the results. Sponsors ought not to establish absolute standards expected to apply to all projects, particularly if resources commensurate with the desired standards of rigor cannot be provided.

3.1.5 Resource Constraints

The concept of an ideal evaluation design is useful to the extent that it establishes standards to strive for. A pilot or field-test design, however, must not overreach the limits imposed by resources. Some of the constraints include:

a) time available

b) budget available

c) type of product

Time available for field testing vocational-technical curricula undertaken in context of a curriculum development project is limited. Not only may a shorter period be available for testing than the curriculum duration but the total length of the project must allow for design, analysis and use of the pilot or field-test results.

Budget available for field testing is generally rather modest. This imposes serious constraints on every aspect of field testing. Most conspicuous is the number of sites indicated for a field test. These are generally specified in advance in Part I projects. The conclusions which may be drawn from a field test are influenced by the number of test sites.

Type of products to be tested and recommended strategies for implementation can have decisive influence on the type of pilot or field test it is sensible to undertake. Some projects produce teacher guides; others include student materials. Some are designed solely for infusion in modest doses into existing curricula.

This latter poses particular concern. It may be difficult to identify the actual use of such a curriculum because it is incorporated flexibly and ad hoc as suitable to the larger course of study.

Even more important is the fact that some curriculum materials designed for infusion do not constitute a cohesive product. This in no way reflects negatively on their value; it may simply be argued that in such a situation the developer does not truly have a complete curriculum but rather only a partial one.

Although in theory anything can be measured, in practice it is often more difficult to measure items of modest scale whose full impact probably depends upon its role in a total program. Thus, for example, an attempt to measure the effectiveness of an occupational cluster curriculum which consists of concepts and content to be selected as appropriate by the teacher probably fails on three grounds:

a) the cost and effort of identifying the specific use of the materials increases the less cohesive the product

b) the effort and feasibility of measuring student outcomes for this kind of product may be unrealistic

c) the importance of student learning in a partial program may be insufficient to warrant certain types of inquiries since a curriculum may be greater than the sum of its parts.

Evaluation methodology must take into acount the total significance of a curriculum in terms of larger educational goals. The techniques of evaluation methodology, because they require definition, distinctions and measurement of units, should not obscure the essential purposes of curriculum improvement or of educational change. Because something may "work" does not guarantee that it is important.

This question is tantamount to asking: are the objectives of a particular curriculum product alone and separate from a larger program significant or likely to produce results? If the answer to this question is negative, assessing the curriculum on operational characteristics only should be considered.

The objectives of a pilot or field test must therefore be:

- attainable within the available time
- attainable within the available budget
- important to answer for the particular products

3.1.6 Pilot and Field-test Objectives: Formulating the Questions

In specifying the objectives of a pilot or field test, resource constraints tend to be the decisive ones. These establish a ceiling on the scale of inquiry the developer may undertake. Within the limit, there are some trade-offs available and decision criteria which may be applied.

Essentially, the choices lie between questions of scope versus depth. While in general it is better to seek answers to questions of effectiveness, in some specific cases (e.g., a partial program implemented in small doses), a large-scale test of usability may be more sensible than a small-scale test of effectiveness. Priorities for the inquiry must be applied by the developer and agreed to by sponsors.

The questions which follow illustrate the kinds of questions which may be answered under different conditions and for different types of products. Most of the questions may be answered in a design restricted to one site, although standards of rigor cannot be met. The developer must ask the highest level question for which he/she can reasonably expect to obtain an answer under the available conditions. Indications of minimum conditions required and maximum necessary are included under scale of inquiry in the charts which follow.

TABLE 3-C.

PILOT AND FIELD TESTS FOR USABILITY

Standards of Rigor

1. All responses to questions of usability have potentially equal validity and require no bases for comparison.

Scale of Inquiry

1. One site or few is sufficient for most questions of usability.

2. Except for certain comparative questions, no more than a few sites need be employed if questions of effectiveness are not also addressed.

Product Characteristics

1. Questions of usability are indicated especially for products with the following characteristics:

   a) highly innovative, untested instructional methods

   b) new content for which no existing department or subject class exists

   c) partial products which do not constitute a cohesive whole or which are implemented in ad hoc fashion in the classroom

   d) curricula designed for implementation by existing school personnel but which are mostly outside the realm of their subject area expertise (e.g., many career exploration curricula, grades 7-12)

User Interest

1. Usability questions are of prime concern to curriculum developers for purposes of improving the curriculum.

2. Some teachers and administrators may be satisfied with results of small-scale tests of usability as the basis for use. These are more likely to be at the individual school level rather than at LEA or SEA level.

Table 3-D.

QUESTIONS OF USABILITY	Scale of Inquiry
I. IS THE PRODUCT USABLE?	One Site
How can the curriculum be improved?	
1. Does a particular curriculum component, strategy or activity work as designed?	
2. Can the total curriculum be implemented in a real classroom setting?	
3. Is the curriculum usable without developer assistance?	
• Are the instructions clear?	
• Are the materials available?	
• Is the reading level appropriate for students?	
4. Can the curriculum or program be installed using normally available resources?	
5. Can a teacher without special training in the content implement the curriculum?	
• What types of training, support, resource personnel are necessary to implement the curriculum?	
6. Do teachers/students accept the curriculum?	
• Is interest demonstrated in the content?	
• Are the instructional methods motivating?	
7. Is the content adequate to meet the objectives?	
8. Are the instructional methods suitable for goals sought?	
II. DIFFERENTIAL USABILITY	
9. Is the product more usable in some situations than in others?	Several-many Sites
• Is the curriculum equally usable in different institutional settings?	
• Can teachers with different training/ backgrounds implement the curriculum equally well?	
10. Can variations in the ability to implement the curriculum be related to sites, student, and/or teacher characteristics?	Many Sites

30

Table 3-E.

PILOT AND FIELD TESTS FOR EFFECTIVENESS

Standards of Rigor

1. The degree of rigor required varies to some extent with the question.

2. A minimum degree of rigor in terms of comparisons of student outcomes is required for statements of learning gains to have any meaning.

Scale of Inquiry

1. Even one site is sufficient to address questions of effectiveness with standards of comparison. The conclusions, however, will be relatively weak given the limitations of a scale.

2. Some questions require large-scale tests or they cannot be meaningfully answered.

Product Characteristics

1. Questions of effectiveness are important to virtually every curriculum test with the following possible exceptions:

   a) partial products which do not constitute a cohesive whole or which are implemented in ad hoc fashion in the classroom

   b) products which are complete but designed for gradual implementation over the full span of public education may require a degree of history or accumulation to be evaluated for effectiveness in a meaningful way

User Interest

1. Effectiveness questions are of concern to all users.

2. Teachers and administrators may be satisfied with results of small-scale tests which demonstrate learning gain as a basis for adoption. These may be found at individual school, LEA and SEA levels.

3. Dissemination standards require degrees of rigor which exceed a one-site test but which may possibly be met with a modest test using multiple sites.

Table 3-F.

QUESTIONS OF EFFECTIVENESS	SCALE OF INQUIRY				
	One Site		Several Sites		Full-Scale Evaluation With Controls
	No Control	Control	No Control	Control	
I. Is the curriculum effective?					
1. Does the curriculum achieve its objectives?	X	X	X	X	X
• To what extent do students demonstrate learning gains in the curriculum content?					
2. Is the curriculum more effective for some objectives than for others?	X	X	X	X	X
• Do students demonstrate greater learning gains in some types of learning than others?					
3. Are the gains substantial and educationally important?	X	X	X	X	X
II. Differential Effectiveness					
4. Was the curriculum more effective for some types of students than for others?	X	X	X	X	X
5. To what extent was the curriculum responsible for measured gains? How did the gains for students receiving the curriculum compare with those for control groups?		X		X	X
6. Was the difference in gains between students receiving the curriculum and control students statistically significant?		X		X	X
7. What kinds of variations occurred among the sites in student achievement gains?			X	X	X
8. How much variation among sites occurred in the relative effectiveness of the curriculum vs. controls?				X	X
9. Is the curriculum likely to be effective when implemented in other sites?					X
10. What characteristics of a site correlate with effectiveness?					X

3.2 Developing the Design

A field-test design refers to the conditions and procedures which guide the collection of data needed to answer the developer's questions. The purpose of the design is to facilitate gathering data which allows the field tester to make valid statements about the outcomes or effects of the curriculum or to organize reporting on its usability.

3.2.1 Working with Givens

With true experiments, there are clearcut methods for design and analysis. With most field tests, however, these often do not apply. Some conditions which are not at the developer's option are known at the outset and constrain design choices. These givens must mold the design. Examples of such constraints include the following concerns:

a) The number of sites is specified, generally at one or three. The level of confidence of the results will therefore be limited.

b) The curriculum itself is targeted at a specialized population of students of whom there are very few for whom control groups cannot feasibly be established.

c) The type of product is designed for differential implementation or must permit implementation decisions by the test sites. The curriculum (the "treatment" in an experimental design) therefore is likely to vary with each site; the program being measured is thus potentially different in each case.

d) The field test necessarily occurs in an action setting which poses problems for evaluation research. In particular, the tendency and even the desire for the curriculum to change as it is being used obscures the developer's ability to identify the treatment or differentiate it from other aspects of the school setting.

Nevertheless, field-test results will be valuable if not conclusive. Even in much larger evaluation activities, results are often ambiguous but they supply hypotheses for further investigation. Field tests of vocational-technical curricula at the current typical levels of fundings may best be termed exploratory inquiries.

3.2.2 Basic Elements of the Design

Three basic elements central to all pilot and field tests must be considered in establishing the pilot or field-test design. The field tester will need to learn from or about at least these three elements:

a) the curriculum itself

b) the students

c) the teacher(s)

3.2.2.1 Curriculum

Since the curriculum is the "treatment" in an evaluation design, it is critical to define as clearly as possible what the treatment is. Otherwise, the field tester does not know what is being evaluated.

In Part I projects, this is a terribly important consideration in the test since many of the curriculum products rely on teacher-elaboration or selection of suggested activities and therefore, the "treatment" itself cannot be known by mere examination of the curriculum guide. This problem arises even where extensive student materials have been developed to support the curriculum objectives.

Four decisions with respect to curriculum must be considered:

a) the definition of the curriculum prior to field test including all required resources or practices considered central to it and the limits of alterations which the developer will accept as still part of the curriculum.

b) a means of documenting what is actually done in the classroom to assess whether the "treatment" was actually the curriculum. Thus, a method for documenting actual instructional strategy must be developed in the planning stage. A classroom log, on which the teacher summarizes the actual use of the curriculum is useful for this purpose. A sample log appears in Table 3-G. A complete set of logs can give the field tester an historical record of how the curriculum was implemented as well as an ongoing commentary on suggested revisions.

c) the importance of adherence to the plan in the course of implementation versus the degree to which changes in the curriculum plan as it is being used may be permitted. The developer must determine the extent to which changes in the plan in the course of implementation will influence his/her ability to evaluate a dynamic "treatment." Some degree of change is not only natural to expect, but it may even be necessary. The acceptable degree of modification depends upon the field-test objectives. Examples of the field-test objectives, associated degree of modification, and probable willingness of school personnel to consent to restrictions on modification are indicated in Table 3-H.

d) the importance/unimportance of uniform implementations of the curriculum by field-test teachers in order to be able to compare outcomes for different classes; or, in other words, the extent to which differential implementation by test sites suits the field-test design.

In all cases the developer's position on these issues must be conveyed to school personnel and consensus of expectations obtained.

TABLE 3-G

Suggested Classroom Log Format

Date_____ Day_____ Teacher_____ Class_____
Concepts:_____
Curriculum Module: Unit_____ Activity Title_____
PP._____

Note: Was this an introductory session to an activity to define terms, etc.

yes no (please circle one)

Method of Teaching: (Discussion, role-play from curriculum, etc.)

Note: Indicate if field visit and type (fill in yes or no).

During Schoool_____ After School_____
With Teacher_____ With other Adult_____
Students only_____
Number of Students per group:_____

Comments on Student Responses:_____

Recommendations on Materials: (e.g. clarity of instructions to teacher, etc.)

Special Note: Indicate (a) differences in teaching method used as compared to curriculum materials:_____

(b) recommendations on alternative presentations:_____

TABLE 3-H

CURRICULUM MODIFICATION

Objectives of Field Test	Degree of Modification Permissible	Likelihood of Consensus by School
Compare outcomes across sites	Low	Low
Identify alternative strategies	High	High
Document Usability	Moderate	Moderate

3.2.2.2 Students

The developer will want information about and/or from the student. It is important to consider that the:

- Type of data varies with the objectives of the field test.
- Extent of data varies with resources of the field test, accessibility of data (confidentiality and school cooperation).

Examples of how field-test questions relate to the type of data necessary appear in Table 3-I below.

TABLE 3-I

DATA ON STUDENTS

	Field-Test Questions	Type of Data
1.	Is the curriculum usable, appealing?	Student reactions
2.	Do students demonstrate achievement of curriculum objectives?	Pre- and post-tests
3.	Is the curriculum more effective for some students than for others?	Background data, including reading level, academic achievement, demographic data for each site, across sites
4.	Specific hypotheses: e.g., "Parental occupations influence students achievement in career and vocational curricula."	e.g., data on parental occupation.

The type of data on or about students which a developer seeks varies greatly and cannot be specified in the abstract. Similarly, access to data will vary with the particular policies of the schools toward making information about students public. Federal legislation now provides for written permission by parents prior to release of student records. Further, the ability of the developer to obtain data depends upon project resources to pay for data collection or alternatively upon the willingness of the school personnel to collect, record and furnish data. Only several general principles are discussed below.

- Relationship of student data to field-test questions: Evaluators often collect data for purposes of "insurance." A clear purpose for all data collected is not always held; the possibility that such information might be useful later is insured by collecting it anyway. In pilot and field-test situations, such insurance is simply too costly for the scale of activity. In fact, the collection of large amounts of data for no clear purpose even in major evaluation activities is unwarranted and has generally proven wasteful.

Data collected on students must be:

- sufficient to answer field-test questions
- relate directly to field-test questions
- serve to answer developer hypotheses

In developing the field-test design, the field-test questions themselves must be prioritized and the developer prepared to omit those questions for which data cannot be secured because of costs, inaccessibility, and/or unavailability of necessary instruments.

- Measuring Student Gain: In virtually all field tests, with the exception of those exclusively formative, the probable minimum data required about students is the extent to which students learn from the curriculum. Student gain generally cannot be assessed unless student knowledge is measured prior to the curriculum and at its conclusion. In vocational-technical curricula, standardized tests or other norms are not generally available. The developer therefore has few resources for establishing student gain besides direct pre- and post-measures of student learning.

Two possible exceptions in vocational-technical curricula may be considered. To the extent that the subject itself is so remote that the developer believes the student is unlikely to begin the curriculum with any basic knowledge, the omission of a pre-test may be acceptable. However, students may learn from non-school experiences (especially at work and at home) about even the most technical fields. Especially to the extent the curriculum teaches career information, initial measures are probably required.

The other exception may be forced by circumstances. For certain types of learning, particularly affective learning, measures may be so difficult to construct within the resource constraints that the developer is limited to a more formative design. This situation may arise especially in career-related curricula for which simple measures of increase in occupational information are obtainable but not meaningful enough for the larger curriculum objectives.

- Access to Data has become an important and controversial current issue. Confidentiality of student records is receiving increasing protection. Explicit written parental approval may be required for student achievement scores and other information. The implications of this are simple: the developer must determine that the data he/she needs on students will in fact be available from the school. The ability of the developer to secure available data is a different question and depends in large part upon resources. This is discussed in Chapter 4. Requirements for OMB approval for certain types of data collection activities is yet a different question discussed in section 3.4 below. In addition, the Privacy Act of 1974 regulates the maintenance of data. Contractors should determine with their project and contracting officers the requirements for maintenance of pilot and field-test data.

3.2.2.3 Teachers

The developer depends to a great extent on information supplied by teachers. The extent and type of information depends upon:

- field-test objectives
- resources

Field tests which focus on usability questions depend almost entirely on teacher input, although most field tests will want a great deal of information on the curriculum from teachers in order to make judgments about it.

Further, the influence of teachers on student learning outcomes can be very great. In order to assess the meaning of the test results, some attention must be paid to teacher characteristics including their knowledge of the subject, method of selection, attitude toward the methods used in the curriculum, etc. This information may be helpful later in accounting for differential student performance. Where field testers seek to learn about differential success in implementation, information about teacher background is required.

Many field tests also require scoring of tests, and collection of other data about students. A major problem for field-test efforts is the amount of data, and consequently, of extra work they require from teachers. Without the data, the developer cannot make judgments but if teacher time requirements are excessive, complete data will not be obtained anyway.

Therefore, reasonable requirements must be established, made absolutely clear, and explicitly consented to by teacher and developer.

3.2.3 Establishing Standards of Comparison

The field tester strives to attain accurate evidence that the observed outcomes are a consequence of the curriculum under investigation. Thus he/she must seek to arrange the conditions of the study so that, at its conclusion, attribution of the outcomes can be reasonably made to the curriculum being tested. The most challenging aspect of the design is the attempt to eliminate or control competing explanations (such as, other student experiences) for the effects of the curriculum.

The documentation of substantial gains for students in one or more sites is evidence that the program is effective in transmitting certain skills, but it is not conclusive. In most situations it might plausibly be objected that the observed gains could in whole or in part have occurred under some alternative conditions. Thus, some estimate of the results which would have occurred under alternative conditions will be needed as a standard of comparison. In a field test situation there are essentially three ways of addressing this problem:

- Randomized control group
- Non-randomized control group
- No control group

3.2.3.1 Randomized Control Group

Randomization is the Rolls Royce of experimental design. In randomization, or random assignment, chance determines which students are assigned to the class(es) with the new curriculum and which to an ordinary class. The advantage of such a procedure is that chance is blind, and hence eminently fair. Thus any characteristic of the students (e.g., age, sex, IQ, eye color) will tend to be equally distributed in the two groups. The groups will be initially identical with respect to all characteristics, whether we have measured them or not. Thus any difference in performance between experimental and control classes is likely to be the result of the only factor on which they differ, namely the program they receive.

To carry out randomization, it is simply necessary to assign students to classes using a random process which gives each student the same chance of getting into each class. For example, suppose that a school has scheduled two classes in a particular subject, and has agreed to use an experimental curriculum in one of them. Then for each student electing to take this subject, a coin could be flipped to determine the class assignment.

The mechanics of randomization are fairly simple and straightforward, and the advantages in terms of an ability to attribute gains to the curriculum are enormous. However, there are several conditions which must be met in order for randomization to be implemented.

- There must be a sufficient number of students available to form at least two classes.
- Complete control over the assignment process must be possible. If classes had to be scheduled so that some students could only take one or the other, for example, random assignment would be impossible.
- It must be possible to maintain the integrity of the design throughout the program. For example, the number of students entering or leaving the program in mid-stream should be negligible.

3.2.3.2 Non-Random Control Group

If random assignment is not possible, the next best thing is to use a non-random control. That is, measure the performance of another class being taught similar material using a standard curriculum. The results can then be compared with those of the experimental class.

The virtue of randomization is that it guarantees similarity of the experimental and control groups, thus allowing outcome differences to be attributed to the new curriculum. With a non-random control group, such differences may be the result of differences between the groups as well as between the programs.

By comparing the two groups on various background characteristics, alternative explanations for the outcome can be explored. Was the class with the new program older, smarter, better prepared? Of particular importance is the pre-test measurement of the skill being taught. A significant difference on this measure prior to the program is a sure tip-off that a comparison of experimental and control outcomes will be biased. Sophisticated statistical techniques can sometimes be useful in dealing with the problem of a priori group differences, but they provide no real substitute for randomization.

In summary then, there are really two conditions necessary for using a non-random control group.

- There must be a sufficient number of students available to form at least two classes.
- It must be possible to assess relevant ways in which the experimental class differs from the control class.

3.2.3.3 No Control Group

Sometimes it will not be possible to have any control group. Data on effectiveness will consist of pre- and post-test scores for students receiving the new curriculum only. What inferences can be made in this situation? The answer is none, unless some information from other sources on what would have happened to the class without the new curriculum is available.

Some curricula may be dealing with material that is so different from anything the students would otherwise learn that any gains can be attributed to the program. A program to teach a highly specialized technical skill may fall in this category. Effectiveness can be measured in absolute terms, since it is likely that without the program, students would be unable to acquire the skill at all. Similarly, an effect might be so dramatic that the only plausible explanation is that it resulted from the new curriculum.

Finally, in some cases, the new curriculum may be designed to replace a standard program for which norms exist based on past experience in the same school and/or other schools. If the relevance of such a norm can be justified, the new program's performance can be compared with this standard.

Unfortunately, it is unlikely that the field tester can approach this rigor. It is often impossible for the tester to randomly assign students or teachers to treatment and control groups. Moreover, control groups equivalent to test groups may not be available because the program under field testing is aimed at a limited number of students. Often project resources simply will not support the operations necessary to establish controls. As a consequence, the designs most likely to be practical for the field tester may be unable to control all threats to internal validity.

The relative importance of these controls must be agreed to by developer and sponsor. Where trade-offs are possible with respect to resources, (e.g., randomization in one site vs. no controls in several sites), priorities may be applied.

3.3 Developing a Data Collection Plan

Data collection is a central activity of any evaluation effort and the selection or development of instruments a major challenge. The design of instruments which accurately assess outcomes poses extremely difficult technical problems.* Their use is equally difficult on technical and non-technical grounds.

For pilot and field-test efforts especially, data collection activities represent the most serious threat to consensus in the entire effort. The kind of information that is critical to the developer may not be of major importance to school personnel. In the data collection process the agenda of the developer and those of practitioners differ most substantially.

No simple solution exists for this problem. To the extent possible, data should be sought of a type and in a form useful to teachers as well as the developer. It is also important to keep in mind that ambitious intentions on the part of both developer and practitioner usually wane as the effort proceeds so that the information is not produced anyway. Realistic time estimates for data collection activities should be made and assessed to minimize the prospect of failure.

A data collection plan must therefore be trimmed to the realities of the situation. The focus of this section is on four basic components of the plan:

- formulating hypotheses
- defining information needs
- selecting a data collection technique
- identifying conditions and constraints

The first and key step of a data collection plan is an articulation of the hypotheses or questions the developer wants to answer. The kind of data necessary follows the choice of questions and implies the method for collecting that data. While time and resource constraints are an ongoing consideration, once the total plan is developed a final assessment of constraints and necessary conditions is essential.

---

* An extensive discussion of instrument design may be found in the literature and is not within the scope of the Guidelines. See, for example, Test and Measurement Kit available by request from Educational Testing Service.

3.3.1 Formulating Hypotheses

The data collection questions establish the boundaries and focus for the plan. Generally, the questions are based on specified program objectives, although often other questions are of central interest to the developer.

The data collection plan assesses different types of outcomes. Commonly developers seek information about:

- student learning or progress
- curriculum manageability
- explanatory evidence

The questions are stated in the form of testable hypotheses. For example, a stated curriculum objective might be, students should be able to demonstrate proficiency in skills relating to travel occupations; specifically, to prepare a flight/travel plan using the Official Airline Guide. In this case, a data collection hypothesis would be: students who receive the curriculum can perform the skills of preparing flight/travel plans and itinerary better than students who do not have the program.

Hypotheses about usability include, for example: teachers will use the curriculum as specified; students will report that the curriculum taught them relevant and useful skills; teachers will report student interest in using curriculum materials; and so forth.

Additional hypotheses might be formulated at the outset about specific explanatory information. For example, general academic achievement or reading level might be an important predictor of student success on outcome measures after receiving the curriculum materials. Or, especially in the case of Part I projects, students' previous work experience or parents' occupations may help to explain differences in success in the curriculum. Such hypotheses would be stated as: success in the curriculum is related to general reading achievement scores; students with prior work experience in any field perform better in the program.

Once the questions have been defined and articulated in the form of testable hypotheses, the next step is to specify precisely what type of information should be gathered that is pertinent to answering those questions.

3.3.2 Defining Information Needs

Earlier sections of this chapter discuss in more detail the types of data and how they vary with the objectives of the field test design. However, at this point it is important to specify how the types of information necessary derive from the field-test hypotheses and relate to the methods of collection.

The type of information, in this case, refers to an operationalization of the field test hypothesis or to defining the category of information required to test the stated hypothesis. Different data collection techniques collect different kinds of information and therefore a statement of the category of information is a necessary step to selecting the appropriate method of data collection.

When this step of defining the category of information is omitted the data collection technique may be chosen inappropriately. For example, collecting teachers opinions about curriculum usability by questionnaire does not test hypotheses about student achievement which requires collection by tests or other measurement instruments.

3.3.3 Selecting a Data Collection Technique

The method of data collection is implied by the type of data specified by the field test hypotheses. Five basic categories of data collection methods are:

- tests and measurement instruments
- observation, including audio and video taping
- questionnaire
- interview
- existing records

An overall understanding of the particular type of data collected by different methods is essential to preparing a data collection plan.

3.3.3.1 Tests and Measurement Instruments

Tests and measurement instruments are used to assess hypotheses relating to student achievement and performance objectives. Tests may be standardized or non-standardized, used as is, adapted or developed specifically for the program.*

Standardized refers to tests that compare students to norms or reference groups, usually nationally representative samples of students in order to assess relative standing. They are based on content and objectives common to many schools throughout the country rather than to specific objectives. Non-standardized or specifically developed tests are not meant to compare students to other groups of students but rather to assess individual progress.

Criterion-referenced tests are a good example of non-standardized tests that are useful in vocational education curriculas. Rather than comparing students to reference groups, they compare each student's own progress on specified performance criteria. They may be more useful to teachers in assessing individual pupil progress as well as to curriculum developers who need to know how students progress through the curriculum.

---

* Certainly the lack of readily available and standardized tests for vocational education is discouraging. However, given the importance of assessing curriculum objectives relating to what students learn from a curriculum, every attempt must be made to collect this type of data.

3.3.3.2 Observation

Observation is most useful to gathering information about student performance of motor skills, curriculum manageability, and contextual information. Observation may be on site classroom observations, tape recordings or videotapes.

Observation is a useful method for assessing curriculum implementation. Knowledge of how curriculum materials were used is important to the developer to explain differences in outcome measures or define the potential flexibility of the curriculum.

3.3.3.3 Interview

Interviews with students and teachers are useful for collecting data about hypotheses relating to curriculum manageability and desirability. Interviews allow flexibility in collecting developmental data for possible revisions.

3.3.3.4 Questionnaires

Questionnaires may range from closed-ended written response to teacher logs or self reports and may be used with teachers, students, or employers. They are most useful to assess attitudes, performance and specific characteristics. For example, developers may hypothesize that students who receive the curriculum will be better prepared for the jobs they enter in that field than students who did not receive the curriculum. A questionnaire to employers would be an appropriate way to collect the necessary data in this case.

3.3.3.5 Existing Records

Existing records, if they relate to the developers initial hypotheses are extremely useful and inexpensive to use. Primarily records offer information on background characteristics that are important for explanation. For example, students with higher reading achievement scores on regularly administered tests scoring consistently higher on curriculum outcome measures might indicate that the reading level of the curriculum is more suitable to higher level readers. Either an adjustment to the reading level could be made or the developer could specify more clearly to the users the prerequisite range necessary to best use the curriculum.

3.3.4 Identifying Conditions and Constraints

In developing a data collection plan, it is necessary to specify who will be responsible for collecting the data, when and how often. These factors are essential to assessing the feasibility of the plan as well as to insuring that the final plan is followed. Such obvious requirements as administering the pre-test prior to the initiation of the curriculum are not always understood and followed by field-test personnel. Once missed, these points in time may be impossible to retrieve. Making them explicit helps to avoid such mishaps.

Obviously time, money and other resources constraints are taken into account throughout the preparation of the data collection plan. Specification of all conditions, such as who collects data and how often, helps to clarify the specific burden on field-test participants to which the test of reason can be applied.

A further use of an explicit and detailed data collection plan is the identification of alternative sources of information and methods of data collection. This serves to cover contingencies, if the plan fails to meet the constraints of reality. For example, if privacy restrictions make it impossible to secure test data on student performance, teacher reports on general student performance might have to be substituted.

The critical factor in data collection is to specify the hypotheses a priori, taking into account real-world constraints. This focusses the effort in order to achieve maximum quality within the boundaries of what is possible.

3.4 Securing OMB Clearance

Pilot and field tests under Part I projects are subject to the requirement for OMB clearance if they involve the collection of the same questions and/or items of information from 10 or more respondents.*

The most accurate source of information about requesting clearance from the Office of Management and Budget is contained in:

- Standard Form No. 83A, Instructions for Requesting OMB Approval Under the Federal Reports Act, and
- Standard Form No. 83, Clearance Request and Notice of Action

Both of these forms may be obtained from the U.S.O.E. Project Officer. This section of the Guidelines is not meant to be a substitute for the information contained in the forms. It briefly discusses clearance procedures and portions of Standard Form 83A.

The Office of Management and Budget provides for the clearance of public reporting and recordkeeping requirements under federally-funded contracts in accordance with the Federal Reports Act in order to minimize the reporting burden of respondents, and to guarantee cost and quality control under federally-supported contracts. For information purposes, it should be noted that under certain circumstances grants also fall within the purview of the clearance process.

The Act obliges Federal agencies to obtain clearance for various documents and plans. In addition, it is the responsibility of these agencies to obtain clearance for documents prepared by their contractors. Responsibility for preparing clearance request forms is shared with the contractor. If there is any question as to whether OMB clearance is required, contractors must obtain a determination from their Project Officer.

3.4.1 Documents Requiring Approval

Standard Form 83A lists documents which require approval, in accordance with the Federal Reports Act. These include:

---

* The Privacy Act of 1974 has implications for the collection and maintenance of data. Regulations associated with the Privacy Act of 1974 are still being developed and Part I Project Directors should stay abreast of these developments through their U.S.O.E. Project Officer.

- report forms, application forms, questionnaires, or interview guides for personal visit or telephone surveys.

- orders, regulations, or other directives which include requirements for respondents to provide information or maintain records to be used or made available for use in the collection of information.

- requests for answers to identical questions which are addressed to respondents in the form of telegraphic inquiries, form letters, information circulars and other devices.

- any supplementary documents involved in these reporting requirements, such as instructions or covering letters.

- plans for data collection**

3.4.2 Material to Be Submitted for Clearance

All documents in the clearance package should be submitted in 8 copies to the O.E. Project Officer. In turn, the cognizant officer will submit the package to the O.E. Forms Clearance Office and in turn to the Office of Management and Budget. The package consists of:

- Standard Form 83, which summarizes type of respondents, data collection method, nature of form, the authority for collection, and so forth

- The Supporting Statement

- All documents to be sent to each respondent, including the report form, instructions, transmittal letter, and any other related documents.

3.4.3 The Supporting Statement

It is of primary importance that the applicant submit a detailed and well developed Supporting Statement. Form 83A discusses the minimum requirements of a supporting statement. Briefly, the following topics must be addressed:

1) Justification,
2) Description of survey plan,
3) Tabulation and publication plans,
4) Time schedule for data collection and publication,
5) Consultations outside the agency
6) Estimation of respondent reporting burden,
7) Sensitive questions, and
8) Estimate of cost to Federal government.

The contractor should not attempt to determine whether OMB clearance is required. Assume it is. Discuss the procedure with the Project Officer. Work with the Project Officer throughout the process.

** From SF 83A, p. 1.

The degree of rigor suitable for or obtainable in a given pilot or field-test design does not influence the requirement for clearance. The degree of rigor does not necessarily dictate the probability of approval for a given submission. OMB criteria for approval take into account the purpose and design of the study. The Supporting Statement is hence enormously important to explain the context and purpose of the study.

BE CERTAIN UNDER ALL CIRCUMSTANCES THAT THE CLEARANCE PACKAGE SUBMITTED IS COMPLETE.

3.4.4 Time Typically Required for Clearance

OMB clearance procedures ordinarily require between six to twelve weeks. Because of the amount of time required, an applicant must plan to submit documents for clearance as early in the project as possible. Incomplete packages are vulnerable to delay.

The schedule for approval means that the complete field-test plan, including the research design, all instruments, etc., must be completed and submitted at least three months before initiation of the pilot or field test. Since field test start-dates often coincide with the opening of the school year, submission dates should be calculated back three months from that point.

3.5 Formulation of Criteria and Policies

A pilot or field-test design is complete only when its technical features and the characteristics of the curriculum have been translated into criteria for site selection and into field-test policies. The specification of both represents the final step in the design which must precede site seeking efforts. Making the requirements of the design explicit helps to insure that they are met and focuses subsequent site seeking.

3.5.1 Criteria for Site Selection

The developer must examine the pilot or field-test design to determine essential characteristics or conditions schools must meet in order to participate in the curriculum test. If, for example, the design requires data on previous student performance, the availability of such data would become a criterion for site selection. A school's confidentiality policies would have to be determined in the site-screening process.

If the curriculum itself required cooperative work experience, and the curriculum test design required comparison to programs without a cooperative element, criteria for site selection would include both cooperative program for the treatment classes and no cooperative program for control groups.

The list below suggests the type of information the developer might need to know about a school in order to implement both the design and the curriculum. Specific characteristics will, of course, vary with each design.

1. Demographic characteristics of the community; of the students in the school.
2. Appropriate course offerings
3. Congruence of experimental curriculum with school and teacher objectives
4. Motivation of teachers to participate
5. Background of teachers.
6. School-community relationships, especially if community field experience is central to the curriculum
7. Number of students: total and per class
8. Method of enrollment of students in classes with respect to randomization; non-random controls
9. Accessibility of student records, other data: policies
10. Availability of release time for test teachers for training; other communications with developers, data collection activities
11. Availability of required funds and willingness to bear costs associated with participating in the field test
12. Policies with respect to student out-of-class activities during school time
13. Willingness of school personnel to perform data collection activities.
14. Willingness of teachers to score and report tests
15. Willingness to permit classroom observation by developer's staff
16. Intent to continue use of curriculum after test
17. Geographic proximity to developer
18. Special equipment needs, if any.

3.5.2 Field Test Policies

The articulation of field-test policies is similarly a part of the design process. In light of design considerations and resource constraints, the developer must translate certain criteria and other implicit operating principles into policy statements which will guide field-test implementation. It is essential that these policies be made explicit before attempting to implement the design. The field tester must be able to make clear to candidate site what is expected of them and be able to screen sites partly on the basis of acceptance of operating policies. Some of these policy issues and their advantages and disadvantages are illustrated below.

Policy Consideration	Advantage	Disadvantage
1. Teacher will/will not be paid an honorarium.	Teachers are more likely to accomplish complete field test implementation	Field test may attract teachers whose interest in money is greater than their interest in the materials; paid teachers may decline to furnish constructive negative criticisms.
2. The developer can/cannot furnish total required number of copies of the materials to the schools.	Field tester is assured that schools have adequate materials to use and are not burdened with time/cost of reproduction	Schools may not be able to afford to duplicate materials. Attitudes towards extra work/cost may affect curriculum.
3. The developer will/will not permit teachers to terminate the test prior to completion.	Developer is able to retrieve some information about curriculum.	School probably will not go along with inflexible commitment to complete.
4. The developer will/will not allow large-scale modifications of implementation plans.	Site seeking is easier since curriculum can be modified by teachers.	Willingness of school will be reduced if curriculum is relatively inflexible.
5. The developer will/will not collect data on students, score tests, etc.	Developer scoring of tests subject to criticism of non-objectivity; but data will be assured	Teacher willingness to collect data varies.

SUMMARY

Goal Setting

1. The objectives of a pilot or field test should correspond to curriculum goals which the field tester wishes to measure.

2. The field tester generally seeks to assess goals of usability and effectiveness.

3. In the hierarchy of curriculum goals, effective student learning is a higher-order goal than usability. Evidence of effective student learning should generally be sought in field-test efforts.

4. The field tester must seek the type of evidence needed by various potential users of the curriculum.

Preliminary Decisions

5. The field tester should make the following initial decisions prior to developing a pilot or field-test design:

   - relevant user decision criteria and standards of evidence
   - suitable curriculum goals to assess
   - type of conclusions sought in the pilot or field test

6. Curriculum product type is an essential consideration in formulating field-test objectives. The less cohesive the product, the more difficult to assess it for effectiveness.

7. Resource constraints of time and money are decisive influences on the design. Initial pilot or field-test objectives must be tested for feasibility against the available resources.

Initial Design Elements

8. Based on the desired conclusions and associated objectives, preliminary pilot and field-test design elements may be specified as follows:

   - information needs and sources
   - frequency and timing of collection of information
   - research conditions (e.g., control groups or not; methods of assignment)
   - numbers of schools, classes, students

9. Of particular importance is the specification of the curriculum and the documentation of its actual implementation, in order to know what is being evaluated.

10. Control groups are generally desirable for curriculum tests which measure effectiveness. Where curriculum content is highly specialized and unlikely to be learned elsewhere, control groups may not be necessary. They are not required for curriculum tests of usability.

11. The evaluation design process proceeds through several modifications based on the application of resource and feasibility constraints. Any aspect of the design is vulnerable to such constraints, including the objectives, the type of conclusions, and the users whom the field tester wishes to satisfy.

Data Collection

12. Based on the field-test objectives, a data collection plan consisting of the following elements should be developed:

   - specify the information items needed and their sources
   - select a data collection technique
   - determine the frequency and schedule for collection of each item
   - identify who should collect the information
   - specify the collection conditions (e.g., testing procedures)
   - select or construct instruments

13. Data collection techniques include observation, interview, questionnaire, testing, and use of existing records. Choice of technique precedes selection of instruments.

14. Responsibilities of site personnel for data collection should be kept within reasonable limits.

15. Criterion-references testing may be particularly suitable for vocational-technical curricula and especially useful to teachers.

16. Office of Management and Budget clearance is required for collection of data from ten or more respondents. ASSUME IT IS REQUIRED FOR PILOT AND FIELD TESTING. The contractor should:

    - assume responsibility for activating the clearance procedure
    - allow three months time for it
    - be certain the clearance package is complete, or delays are likely to be encountered
    - give sufficient attention to the Supporting Statement
    - work with and submit the package to the Project Officer who will submit it in turn to the O.E. Forms Clearance Office, for transmittal to OMB.

Criteria and Policies

17. The design process is complete only when curriculum characteristics and the pilot or field-test design are translated into explicit criteria for site selection and operating policies for the curriculum test.

Chapter 4. Site Seeking, Finding, Holding

4.1 Overview of the Process

The process of selecting sites for a pilot or field test involves identifying candidates, communicating with candidate sites and reaching consensus with selected sites on all aspects of the test.

Site seeking, finding and holding establishes the arena for the pilot or field test and requires the same care as that devoted to the research design. In this process, the curriculum developer must provide and obtain information sufficient for informed decision-making by school personnel and the developer. This means that the developer should take responsibility for providing information even beyond that which administrators or teachers may solicit to assure that all parties are fully aware of the requirements and limitations of the pilot or field test.

Securing pilot or field-test sites may be thought of as a negotiation of interests over a period of time, culminating in an agreement to cooperate or to disengage. Early information-sharing can condense the time required for negative screening.

4.2 Identifying Potential Candidates

The developer is interested in identifying those sites which meet his/her criteria for pilot or field testing of the materials. Contacts through previous associations or previous letters of inquiry regarding the project are good sources of potential candidates. Identifying a school through a mutual associate or previous association is a superior method to a cold call to a school system where the developer is unknown. For the developer who lacks relationships with individual school sites, there are several strategies for identifying schools which may be appropriate for pilot or field testing.

1. Solicit suggestions from the Project Advisory Panel and consultants.

2. Contact State Department of Education personnel with responsibilities in the field related to the curriculum. For example, the State Curriculum Supervisor in the relevant vocational area may be able to provide suggestions for sites that would be interested in pilot or field testing curriculum materials.

3. Contact the officials of professional associations or unions in the field related to the curriculum and solicit recommendations concerning suitable sites. For example, a state council of the National Association of Industry-Education Cooperation may be helpful in identifying sites which are interested in pilot testing a model for a student placement service.

4. Contact officials of student clubs in the field related to the curriculum and solicit their recommendations concerning suitable sites. For example, officials of the Future Farmers of America may

be helpful in identifying sites which are interested in field testing a curriculum of Agribusiness.

5. Solicit suggestions from State and Federal sponsors and from the directors of the national network of curriculum coordination centers for vocational education.

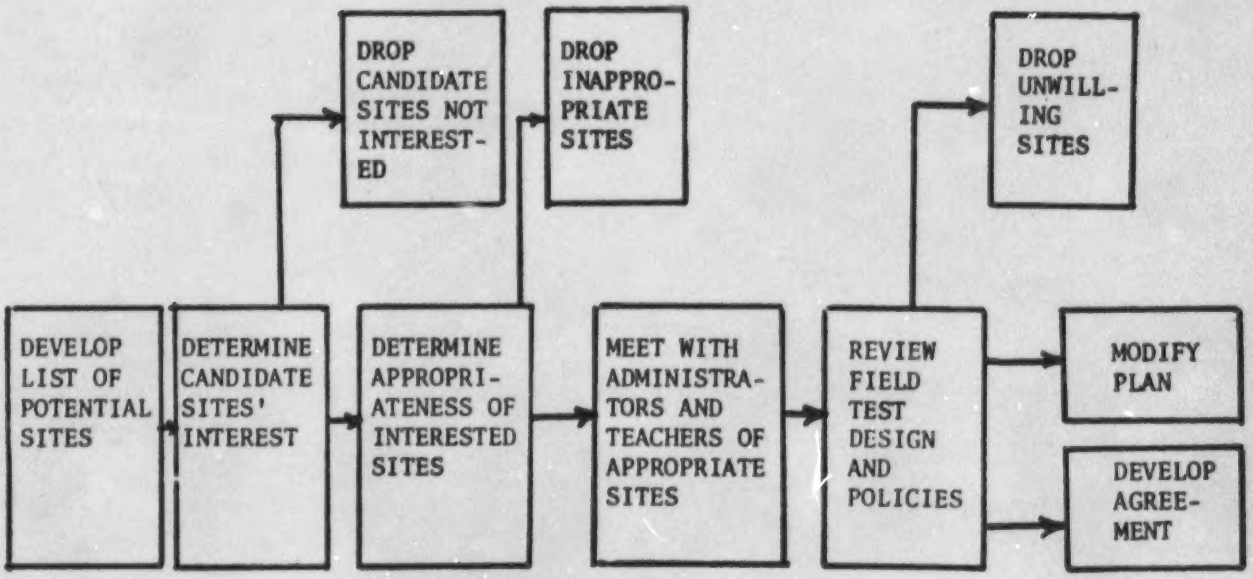
4.3 Contacting Potential Sites

Once the developer has a list of potential sites, it is necessary to communicate directly with the sites to determine their interest and suitability to the field test. The list of potential sites will most likely include a mixture of teacher and administrative contacts. Although it will eventually be necessary to involve both teachers and administrators, the developer should first contact the individual to whom he has been referred. If this person is interested, he can later assist in identifying both the personnel whose approval will be necessary and those personnel who will need to be involved. It will eventually be mandatory to involve all personnel.

Communications with potential sites can be thought of as information exchanges necessary for decision-making by both parties. Careful and continual screening of a site's interest in and appropriateness for the pilot or field-test design helps to weed out inappropriate candidates at the earliest points. Table 4-A which follows summarizes this process.

TABLE 4-A

SITE SELECTION PROCESS


DROP CANDIDATE SITES NOT INTERESTED
DROP INAPPROPRIATE SITES
DROP UNWILLING SITES
DEVELOP LIST OF POTENTIAL SITES
DETERMINE CANDIDATE SITES' INTEREST
DETERMINE APPROPRIATENESS OF INTERESTED SITES
MEET WITH ADMINISTRATORS AND TEACHERS OF APPROPRIATE SITES
REVIEW FIELD TEST DESIGN AND POLICIES
MODIFY PLAN
DEVELOP AGREEMENT

The outline below describes the information exchange which should take place between the developer and the site in order for both parties to be able to decide whether to participate together or disengage.

1. Information about the Developer

   - Nature of the organization and project staff
   - Previous experience

2. Description of the Project

   - Funding agency
   - Objectives and characteristics of the curriculum
   - Target population: grade level, subject class
   - Duration of the curriculum

3. Description of the Field Test

   - Purpose of the field test
   - Types of teachers required
   - Advantages to the school of participating
   - Number of schools/classes participating (specify other school systems already committed, if any)
   - Schedule and duration of field test
   - Costs, if any, to participating schools
   - Research aspects of the pilot or field test
   - Responsibilities of the developer
   - Responsibilites of site personnel

4. Information about the School System

   - Course offerings relevant to pilot or field test
   - Teacher backgrounds and other personnel necessary for pilot or field test
   - Student enrollment
   - Program characteristics (e.g., cooperative programs) relevant to pilot or field test
   - Enrollment procedures
   - Community relationships, if relevant

5. Information about Site's Ability to Meet Field-Test Requirements

   - Availability to staff for training
   - Willingness to maintain curriculum logs
   - Willingness to score tests
   - Availability of student data (policies regarding confidentiality)
   - Ability of students to participate in field trips, if required (including ability to pay for transportation; parent's approval, etc.)
   - Willingness of teachers to conform to the curriculum plan

In communicating with schools, it is useful to make explicit the advantages of participation to the school. A successful curriculum test is more likely if the school's agenda is served by it. In communicating with sites, the developer should:

1. Make certain that all correspondence is personalized and reflects a professional understanding of existing local, regional and state priorities.

2. Attempt to relate how proposed project's goals relate to the school district's policies and operations.

3. Demonstrate clearly the advantages of local participation.

4. Identify exactly the extent of staff and/or student involvement.

5. Explain the nature of the services to be rendered and categorize the human and material resources that will be provided and required.

6. Specify the anticipated outcomes that can be produced for immediate and long-range benefits.

7. Emphasize the public relations value of the proposed project activities.

The developer will discover that the level of detail increases as the communications proceed. The initial contacts with potential sites involve determining the site's interest on the one hand, and suitability in terms of developer criteria on the other. Because of the considerable, costly effort which can be spent in selecting sites, the developer should pre-screen sites on the basis of the fullest possible information on the telephone and by mail prior to actually meeting with school personnel. Clearly, school personnel unwilling or unable to fulfill the requirements of the field test should be dropped from consideration or in the absence of an alternative, the requirements of the field-test design may have to be modified.

Subsequent contacts with appropriate potential sites will involve more discussion of the obligations of the site and the developer. It is crucial at this stage to involve all of the actors. If previous contacts were with administrators, the teachers must now be involved. Securing the interest and approval of the person in authority is insufficient to assure full participation. The teachers and sometimes the students who will be involved need to be willing to participate, and direct communication with them is a prudent method of reviewing candidate schools.

Likewise, if previous contact was with teachers, the administrators must now be involved. Their support is essential to a successful effort. Since the field test usually involves additional effort beyond their usual duties on the part of the site staff, it is necessary to involve all those who will undertake the additional effort required to participate in the field test.

Prior to final selection of the site, it is imperative to review exactly what is required of the administrators, teachers, other staff such as counselors,

the students, and perhaps the parents. So that participants may enter a field test fully knowledgeable about what the commitment entails, issues such as the following must be fully understood:

- Costs, if any to participating sites (e.g., for reproduction of materials conduct of field trips, etc.)
- Data requirements, willingness of school to make it available and safeguards to privacy
- Identification of responsible parties for data collection, including willingness of teachers to score pre- and post-tests; maintain curriculum logs
- Ability of students to participate in field trips, if required (including ability to pay for transportation; parent's approval, etc.)
- Degree of cooperation which may be expected from out-of-school personnel, such as local businessmen, if important to implementation of the curriculum
- Degree of flexibility teachers will have in the implementation of the curriculum

With respect to the latter issue, a useful policy is to permit deviation from a plan on a discussed, negotiated basis in the instance. That is, teachers must inform the developer of the need and/or desire to modify the plan and both must consent.

The number and nature of communications between the developer and the sites varies with their relationship and proximity. A typical pattern between unfamiliar parties is as follows:

1. Initial Contact: generally by telephone, and focusing on descriptive information about the developer, the curriculum, the field test and the school system.
2. Follow-Up Contact(s): generally by mail, documenting the information conveyed by telephone, and including a sample of the curriculum materials or plan. For schools local to the developer, follow-up contacts may be accomplished in person. Requirements which the school must fulfill to participate in the pilot or field test should be clearly spelled out.
3. Clarification and Decision Contact(s): For schools interested and eligible after the requirements have been discussed, all school personnel potentially involved in the pilot or field test should participate in review and discussions of all aspects of the curriculum and field test and the respective responsibilities of all individuals involved. If possible, these discussions should be held in a face-to-face meeting. Where travel budget limitations do not permit a meeting, extensive phone conversations with teachers and administra-

tors should be held. All actors must then agree to assume the designated responsibilities, or agree to disengage.

4. Confirmation of Decision: The decision to participate or not should be confirmed in writing. At any point that a school drops out or is dropped, a thank you letter should be sent. For positive decisions, the mutual understanding of all responsibilities should be detailed in writing.

Early planning of pilot or field testing is essential to assure adequate lead time to identify alternative sites should negotiations with initial candidates prove unpromising. Modifying the test design should be considered as a last resort, and every effort should be made to locate sites which meet the requirements of the field test.

Field testing with sites which cannot be monitored by the developer is generally not recommended, because misuse or ineffectiveness of the untested curriculum can lead to poor acceptance of it later. Further, little or no useful information accrues to the developer, although expenses may be incurred (e.g., in furnishing the materials to a school). However, in cases where the developer is primarily interested in a teacher's ability to use the materials unaided, unmonitored sites may be a suitable addition.

4.4 Reaching Consensus

Once the developer has determined that the candidate school understands and will fulfill the field-test requirements, it is a good policy to record in writing the mutually agreed-upon responsibilities of the developer and the site. This written document is not meant to be a formal contract to bind the developer and the school, but is designed to ensure that both the developer and the school understand their responsibilities in the pilot or field test. The written document should synthesize the previous discussions of the curriculum developer with the site personnel and indicate responsibilities for accomplishing the curriculum test plan in line with the overall test objectives. Potential items for inclusion in the written document are identified in the summary which follows.

Memorandum of Agreement

Duration of Curriculum: length of curriculum test

Student Selection Procedures: especially when research design requires specific procedures.

Training: the amount of training which the developer would provide to teachers and others, both in the curriculum materials and in instrument use.

Assistance and Monitoring: the amount of technical assistance provided by the developer and its form (e.g., site visitations, mail communications, or telephone conversations) and the availability of site personnel.

Materials: the quantity of curriculum materials, reporting forms, test instruments, etc., furnished by the developer and reproduction requirements of the test site. Party responsible for reproduction of materials should also be indicated.

Data Collection: the data required by the developer, party responsible for its collection, such as demographic information, correction of pre- and post-tests.

In order to accomplish data analysis on schedule, specific dates for accomplishment of data collection tasks and transmittal of data should be specified.

Implementation Plan: site's schedule and form of curriculum implementation and procedures necessary to modify plan by mutual consent.

Teacher Records: methods for reporting curriculum recommendations.

Reports: the developer and test site's obligations for written or oral progress reports should be specified.

Evaluation: requirements of the developer in reporting out on the field test and reporting back to the site.

Dissemination: developer's responsibility for furnishing test site with revised curriculum materials following completion of field test and revisions.

SUMMARY

1. Making congruent the expectations and objectives of all participants in a pilot or field test is critically important to a successful effort.

2. All school personnel involved in the pilot or fielt test, including adminstrators, teachers, guidance counselors, must be fully informed about their responsibilities and must consent to them.

3. The site development process required determining whether candidate sites are interested, meet design criteria, and can fulfill the obligations of the pilot or field test. Sites which fail to meet the requirement should be dropped.

4. The process is essentially one of information exchange about the developer, the curriculum, the field test, the school system, and the site's ability to meet requirements of the test, culminating in a mutual agreement to participate or disengage.

5. The fullest possible information should be provided to schools as early as possible to reduce the time and cost of site selection.

6. Identifying potential sites through known contacts is more promising than a "cold" call.

7. Initial communications are primarily by telephone and mail and focus on descriptive information about both parties and the curriculum. These are followed by discussions of the obligations of each party and the identification of specific teachers and classes for the curriculum test. Final decisions by both parties are best made after an in-person meeting of all affected personnel.

8. A memorandum of agreement is a useful written tool to record the mutual understanding of the developer and the site. It should cover all aspects of the curriculum test and the responsibilities of all participants in it.

9. Early planning allows the developer time to assess alternative candidates, drop inappropriate sites, and maintain the integrity of the evaluation design.

10. The developer should support the site's participation by furnishing necessary or desirable letters to school officials; publicity releases to the local press; etc.

Chapter 5. Implementation of the Pilot or Field Test: Interaction Between the Developer and the Site

The relationship between the developer and the site(s) during the implementation of the pilot or field test depends upon the nature of the evaluation design and the resources available for accomplishing it as reflected in the memorandum of agreement.

Implementation of the pilot or field test involves the following four major processes:

1. Teacher training
2. On-site technical assistance and support
3. Assessment of implementation
4. Data collection, including testing of students

Some of the processes above--training teachers, technical assistance--represent interventions by the field tester which may or may not be suitable for the field-test objectives. For example, the developer who wishes to test the usability of a curriculum guide may elect not to provide technical assistance during implementation. Nevertheless, the kind of information the developer needs to improve the curriculum usually makes it desirable to interact with the teachers, although resource constraints typically keep this within a limited scope.

The interactions between the developer and test site can be divided into three major types related to phases of implementation: pre- during, and post-field test. The purposes of these interactions during each of these phases are discussed below. These patterns are illustrative and must be tailored to the research and resource requirements of each project.

5.1 Pre-Implementation Activities: Teacher Preparation

The essential purpose of pre-implementation activities is to prepare teachers for their role in the pilot or field test.

Teacher training functions may be accomplished by visiting each school or by bringing teachers, or a sample of them, to a central meeting place. If possible, it is often useful to bring teachers together to one meeting place for initial training and efficient planning. Particularly if the field-test design dictates conformity in implementation plans, such a meeting is especially efficient.

Where resources are limited, one teacher from a site with multiple test classes might be brought to a central meeting and designated to coordinate planning, preparation and training of other teachers at the site. The developer may verify the comprehensiveness of the training performed by the representative through a check on the implementation plans and associated schedules developed by the other teachers.

Where meetings simply are not feasible, an attempt must be made to plan by substantially less desirable methods of detailed written instructions and follow-up phone calls. More local sites might be considered if travel costs represent the major barriers to pre-implementation planning.

Whatever strategy is employed for teacher preparation, the following objectives should be met:

- training in the curriculum
- training in the field-test design
- development of implementation plans
- scheduling of future contacts

Each of these is discussed in turn below.

1. Training Teachers in the Curriculum

   Teachers should be trained in the objectives, instructional strategies, classroom organization and/or implementation requirements (e.g., cooperative community relationships) of the curriculum.

   If teachers are not familiar with the content or subject matter of the curriculum, some training may need to be provided here. References for teacher preparation should be identified.

   It is advisable to provide instructors with parts of the curriculum package as early as possible to enable them to review the content and methods, to order required materials, etc. If the materials are not yet complete, alert teachers to those planning requirements which must be undertaken early.

   Particularly where teachers must elaborate curriculum outlines, sufficient time must be provided.

2. Train Teachers in the Evaluation Design

   The objectives and methods of the pilot or field test should be explained and identification of other participating schools made. The importance of the teacher to effective completion should be reiterated.

   The developer should train teachers in procedures for data collection and in the proper use of recording or reporting forms. This includes procedures for administration and scoring of pre-tests and interim tests, and methods of curriculum documentation, such as the log suggested in Chapter 3 (see Table 3-G, p. 27).

3. Supervise Teachers in the Development and Implementation Plans

   Teachers need to construct their own plans for implementation of the curriculum materials which delineate the portions to be taught, the amount of time the teacher intends to spend on each exercise or

unit, and the real-time schedule for doing so. The plan should actually indicate which exercise/unit will be covered on each class day. On the plan, special notes should be made which describe strategies for infusion (if that approach is being used); other subject areas which will be covered; those portions of the curriculum which will not be used; and modifications of the developer's materials.

The plan is formulated in considerable detail to insure that the instructor will be able to accomplish implementation; provide evidence of the teacher's familiarity with the materials and enable the developer to prepare a monitoring plan in relation to implementation. Written copies of the implementation plan should be shared by teachers and developer. An excerpt from a teacher's implementation plan appears in Table 5-A below. A simple calendar will serve.

A key problem which may arise in the initial planning activities relates to the discovery from teacher implementation plans or other sources that a teacher does not understand the curriculum or intends to use it beyond limits acceptable to the developer. Despite careful screening and negotation discussed in Chapter 4, above, such instances arise. The developer should consider dropping the teacher from the test plan, furnishing additional assistance, or reviewing the initial commitments in hopes of restoring it.

TABLE 5-A

SAMPLE TEACHER IMPLEMENTATION PLAN

DATE	EXERCISE/UNIT	PAGE NO.	ADDITIONAL INFORMATION
12/19	Preparing for the Interview; Resumes	7-24	Homework: Personal Data Sheet for notebook
1/6/75	Preparing for the Interview/ Application Forms	7-28	Film
1/7	Job Characteristics	7-32	
1/8	Attending the Interview Preparation for Role Playing	7-34	Homework: Assign roles to be studied

4. Schedule Visits, Phone Calls, Due Dates for Submission of Data

The developer and instructors should confirm a schedule for completion of various data collection tasks and transmittal dates to the developer.

Remember that a teacher's day is often fully obligated. Schedules of subsequent visits, if any, and phone calls should be prepared in light of the school calendar and teacher's daily schedule of classes and free periods.

TABLE 5-B

TENTATIVE SCHEDULE FOR DATA COLLECTION	SITE: _____
TASK	DUE
CORRECTION OF PRE-TEST	October 1
COLLECTION OF ALL DEMOGRAPHIC DATA	November 1
PROGRESS REPORTS	Fifteenth of Each Month
CORRECTION OF POST-TESTS	February 1

Tentative Visitation Schedule

DATE	TIME	SITE	TEACHER	EXERCISE
9/20	9:00	Franconia	Bullock	Social Services
9/21	11:00	Franconia	Weinstein	Economic Services
9/22	8:00	Albertville	Frank	Public Safety
9/23	8:00	Manchester	Georgi	Corrections

5.2 On-Site Monitoring

Observing the curriculum in action serves two important purposes:

- it provides the developer with the basis for assessing implementation, interpreting field test results and planning revisions through insights gained by direct classroom observations
- it provides school personnel with the benefit of assistance and support from the developer, which may be one reason for their participation.

During visits, the developer can observe the curriculum in operation, assess student reactions, and evaluate teacher implementation of the curriculum. Such observations provide a context for interpreting teacher reports and can verify or modify them.

The developer can answer teacher questions and discuss teacher and student recommendations and judgments. On-site visits also open the opportunity for developers to communicate with administrators in support of the field-test activities. Teachers and administrators appreciate such evidence of developer commitment to the field test. It is useful to prepare brief written reports after communication with teachers in order to document their comments and observe reactions, and to communicate progress to the project sponsor.

The developer must be cautious in drawing conclusions based on observations of a single or even several classes because of the special conditions which an observer introduces and the potentially unrepresentative nature of brief, periodic visits. Scheduling visits to observe the same curriculum units in different sites furnishes comparative information on different applications or degrees of success. This may help to identify needed revisions where common problems arise across sties.

A number of housekeeping functions may also be accomplished during on-site visits, including collection of test results, curriculum logs, etc. Insight into data collection procedures may also be obtained and this is crucial to interpreting results.

If budget permits, monthly visits to each site are useful. Where impossible, the developer may hire a local affiliate, or designate one site teacher to report on implementation and collect data. Such approaches may be acceptable if a working relationship with the developer already exists.

Some consideration should be given to substitution of local sites for distant ones to permit opportunities for observation even within limited budgets. If all else fails, pre-scheduled, periodic telephone contact may be useful to expand contact between developer and site personnel. However, it is a pale though sometimes necessary substitute for site visits.

A sample guide for site monitoring follows:

TABLE 5-C

GUIDE FOR SITE MONITORING

1. Assess General Student Reaction
   A. From discussion with teacher
   B. From class observation

2. Assess General Teacher Reaction

3. Observe Class
   A. Summary of class activities

4. Review and Collect Classroom Log Sheets

5. Review/Collect Results of Pre-Tests (Test and Control Groups)

6. Review/Collect Class Demographic Data, Interim Tests

7. Discuss Level of Difficulty of Curriculum
   A. Discuss problems
   B. Social interaction of slow and advanced student
   C. Methods of dealing with problems

8. Determine Use of Outside Resource Personnel and Field Trips

9. Discuss/Decide Any Modifications to Original Implementation Plan

The best planning will not guarantee against disrupted visits. The developer may discover that the class to be observed has been cancelled due to an assembly or that the test teacher falls ill when he arrives on his second visit of three from 1000 miles away! These are normal events in the process which cannot be totally prevented.

Attrition problems are most serious. In the course of conducting a curriculum test, one problem which often arises mid-stream is the failure of a teacher to follow a plan, to collect and record data, etc. Proper monitoring and technical assistance might circumvent this problem, but not always. Generally, the problem arises too late to replace the class. Given the usually small samples involved in the testing of curriculum projects, the attrition of even one class can reduce seriously the developer's ability to make statements about the curriculum. Constructive learning, however, may be derived from attrition when the developer attempts to determine why the curriculum is not being fully implemented. The developer can learn about the limits of applicability and likely misuses of the curriculum. Based on this assessment, the curriculum can be modified accordingly.

5.3 Post-Implementation: Closing the Loop

The final phase of pilot or field-test implementation involves:

- obtaining the overall recommendations of field-test personnel
- arranging for post-testing of students
- closing the loop with the sites

Immediately following or near to the completion of the curriculum, teachers are best able to reflect on and summarize their experiences with the total curriculum. The developer should plan a lengthy debriefing to solicit judgments about the total curriculum, recommendations for format, revisions, additions, relevance, parental and administrative reactions. Some important judgmental information which may not have been foreseen or simply cannot be recorded on a form may be obtained in this way.

Since the final visit is likely to occur prior to post-testing of students, developers should also clarify procedures for administering it. Arrangements for transmission of post-tests should be made.

Formal letters of appreciation to administrative and instructional personnel should be sent. Reports on completion of the pilot or field test to local newspapers, parent groups, etc., should be offered.

The developer should plan to report field test results to site personnel; and the revised curriculum should be furnished when it is complete.

SUMMARY

1. Implementation of the pilot or field test involves teacher training, on-site assistance and support of school personnel, observation of the curriculum in action, and data collection functions.

2. Teacher preparation prior to initiation of the field-test includes:

- delivering the materials
- training of teachers in the curriculum
- training of teachers in the evaluation design and overall specific data collection procedures
- supervision of teacher-developed implementation plans
- pre-scheduling of meetings, visits, telephone calls.

3. Teacher preparation functions are best accomplished in person. Budget constraints suggest choosing sites local to the developer.

4. On-site monitoring is important in order to:

- assess the way in which the curriculum is actually implemented
- gain information from observation and discussions with teachers and students for interpretation and planning of revisions
- provide school personnel with assistance and support.

5. The developer must be prepared for imperfections in the monitoring plan, such as disrupted visits; poor implementation; defects in the data collection procedures, schedule or completion.

6. The final phase of the field test offers an opportunity to obtain the overall judgments and recommendations of teachers.

7. Closing the loop with the schools is an important function and includes:

- formal letters of appreciation
- notification of field-test results
- provision of the revised curriculum package when complete
- publicity releases for the sites, if appropriate

Chapter 6. Data Analysis and Interpretation

Analysis and interpretation of data offer the field tester the prospect of answers to his/her initial questions. Analysis as used here refers to the specific techniques required to convert raw scores and other data into aggregate, reportable form. Interpretation of data here refers to making sense out of the aggregate data from the perspective of drawing conclusions about the curriculum as a whole. In pilot and field test efforts, interpretation of results poses a major challenge. With small sample sizes, newly-developed instruments, data gaps, etc. making sense out of the information obtained will be difficult.

Since standards of rigor are often unattainable in field-test efforts, it is possible that no conclusions can be drawn. Responsible use of the data is essential; no conclusions can be fabricated where they do not exist. A simple unelaborated conclusion that no effects could be measured, however, would rob the potential users of important contextual data, including the insights of the actors involved. While intuition is insufficient basis for decision-making about a curriculum, it is important information for directing future exploratory efforts.

Even where positive evidence of effectiveness is produced, evaluators must consider the broader context of a new curriculum and assess the suitability and importance of the program for educational improvement. Reporting statistical data with precision is only part of the job

Extensive consideration of analysis techniques are beyond the scope of the Guidelines. Two basic useful techniques, however, are discussed below. These are:

- simple comparison of scores for treatment and control groups
- use of item analysis for vocational-technical curricula

6.1 Comparing Experimental and Control Outcomes

Data resulting from a typical field test may include pre-test and post-test measurements on all students receiving either the experimental curriculum or a standard (control) curriculum.

To make definitive inferences about the curriculum's effectiveness requires a comparison between experimental and control group performance. The basic data for answering this question consists of the four sets of scores corresponding to the experimental pre- and post-scores, and the control pre- and post-scores. As a first step in the data analysis, summary statistics such as means, medians, and standard deviations can be examined. For most purposes, the group means are the best summary measures, and the pattern of these four means form the basis for inference on program effectiveness.

The pattern of means can be represented in terms of three characteristics:

- the pre-test difference between groups,
- the gain score difference in groups (post minus pre),
- the post-test differences.

Eleven distinct situations may arise. These are illustrated in Table 6A.

In the first case, there is an initial gap between the groups which becomes wider after the curriculum. There are two principal interpretations which must be considered. One is that since the experimentals gained more than the controls, the experimental program is superior. A second interpretation is that the initial difference resulted from differential growth rates for the two groups. In the case, we would expect the gap to widen even if the new curriculum was not better. Judgment must be applied to decide whether the size of the apparent effect could be plausibly attributed to differential growth. Case (2) represents the same situation with the roles of experimental and control groups reversed.

Table 6A

Experimental Minus Control

	Pre-test	Gain	Post-test	
1)	+	+	+	
2)	-	-	-	
3)	+	-	+	
4)	-	+	-	
5)	+	-	-	
6)	-	+	+	
7)	+	0	+	
8)	-	0	-	
9)	0	+	+	
10)	0	-	-	
11)	0	0	0	

E = Experimental Group

C = Control Group

Case (3) represents a situation where the initial difference in favor of the experimental group becomes smaller at post-test time. The obvious inference is that the experimentals have not done as well as the controls. On the other hand, it is possible that the nature of the subject matter or the test creates a "ceiling effect." In a 20-item test, for example, some of the experimentals may have reached the maximum score of 20. Such a situation must be ruled out before declaring the new curriculum a failure. Similar considerations apply in case (4) with the roles of experimental and control groups reversed.

Case (5) involves the clearest interpretation. The control group starts out below the experimental group and ends up above. In the absence of strong evidence to the contrary, one would conclude that the experimental curriculum is ineffective. Similarly, in case (6) one is likely to conclude that the new curriculum appears to be effective.

In cases (7) and (8) the obvious inference is that the experimental curriculum is as effective as the control. Possible differential growth rates and ceiling effects should, however, also be considered in interpretating these results.

Finally, in cases (9), (10) and (11) the experimental and control classes begin at (approximately) equal levels. Then any post-test difference may plausibly be attributed to the different curricula. Under randomization, this situation is very likely, though it may occur in other ways.

To the patterns of mean scores above, now add numerical values. On a particular test the following means are observed:

	Pre	Post	Gain
Experimental	6.4	16.8	10.4
Control	6.2	12.2	6.0

It seems reasonable to infer a positive effect of new program. To make this inference more precise, a statistical test of the significance of the observed difference in gain scores (10.4 - 6.0 = 4.4) is applied. This test determines the probability that a difference this large would have arisen by chance even if there really were no difference in effectiveness between the two curricula.

Statistical significance provides evidence on whether an observed effect is real, but gives no information on whether it is important. The determination of the educational significance of an effort must be somewhat subjective. In terms of the particular test used, the judgment must be made whether an improvement of 4.4 points is educationally significant.

While the primary analysis rests on test scores, other information collected on students can also be valuable in interpreting results. Whether or not randomization is used, it is valuable to record the means on background variables which may be related to outcome. Particularly when classes are small there may be an imbalance between experimental and control groups on some

important variable. Not only can such imbalances cause pre-test differences, but a program might be more effective for some kinds of students than others. For example, students with previous work experience might perform better in vocational curricula involving a cooperative component. If these students were overrepresented in the experimental class, attribution of learning gains to the experimental curriculum would require modification.

6.2 Item Analysis

Comparisons of total scores on an entire test is convenient. However, it may sometimes obscure subtle effects involving specific skills. Thus, when possible, the test can be broken down into sub-parts, and even individual items. The percentage of students answering any item correctly can be used in the kind of analysis described above. Individual items have much less reliability than larger tests, but provide much more detail on exactly what is and is not being learned.

Item analysis is a promising strategy for the type of curricula developed under Part I. These tend to include a mix of occupational competences, career decision-making, skill development, and occupational information. A curriculum may have differential success in achieving these objectives, and assessing student scores on different clusters of items may be instructive to the developer. For example, the developer may learn that a curriculum does not appear to train students in management skills as effectively as in occupational information. Plans for revising the curriculum may be better targeted to selected portions based on this type of information.

6.3 A Practical Consideration

The small number of students, classes and schools in a pilot or field test weakens the analyst's ability to generalize findings about curriculum effectiveness. The problem of small sample sizes is often complicated further by loss of data. Teachers sometimes fail to administer or to score student tests. From the perspective of data analysis, three test classes may be reduced to only one if the data is not supplied.

Any attempt to collect excess data for "insurance" purposes will not solve this problem. In fact, the increased burden may result in <u>less</u> data. Generally, the larger the teacher's responsibility for data collection, the greater the risk that not all data will be furnished. Such situations arise even where responsibilities for data collection are willingly assumed and consented to.

It may be best, therefore, to deploy developer staff to collect data and minimize teacher responsibilities in this area. This strategy is feasible if two important conditions are met:

- project resources must be adequate to support this undertaking
- tests administered to assess student learning are objective, if scored by the developer

If the resources necessary for project staff to collect data exceed those available, reliance on school personnel becomes a typical alternative. If performance tests require judgment and interpretation, the developer's objectivity and credibility may be questioned. Adequate resources and careful choice of test design (e.g., machine scorable) could eliminate the reasons for reliance on school personnel for data collection. Productive learning from pilot and field tests could thereby be enhanced.

SUMMARY

1. Simple statistical reporting of field-test results must be elaborated with interpretations and contextual information. This is particularly important where methodological constraints (e.g., small sample size) do not permit generalization from the data.

2. Educational significance of reported gains should be assessed. Scores alone do not signify the importance of reported learning achievement.

3. Alternative explanations for differences in pre/post-scores and between experimental and control groups must be explored in interpreting results. Data obtained from observations, teacher reports, etc., help to interpret outcomes.

4. Item analysis is a useful technique for determining differential effectiveness of curricula particularly if they have several different types of learning objectives.

5. Imperfect and incomplete data collection by site personnel constrains the developer's ability to analyze and interpret the field test.

6. Deployment of developer staff, instead of school personnel, to collect data may improve the fulfillment of data collection tasks, if adequate resources are made available for doing so.

Chapter 7. Use of Results

Whether the results of the pilot or field test are positive or negative, the developer should be responsible for:

- conveying the results in suitable form to interested audiences
- making explicit the limitations of the findings and the design

The specific results of the pilot or field test are important in shaping:

- plans for curriculum revisions
- dissemination decisions

The findings of the curriculum test, however, do not always offer clearcut and positive direction for revision and dissemination activities. The developer must responsibly confront ambiguous or negative findings and exercise the necessary judgment to act on the evidence available. In the case of curriculum revisions, a variety of sources may be brought to bear. Dissemination decisions, however, to which curriculum revisions ultimately are linked, are more difficult to make in the absence of positive field test results. Dissemination is also extremely difficult to reject because of the investment of resources made by the time the pilot or field test is completed. Bases for planning revisions and considerations for effective dissemination are discussed in turn below.

7.1 Curriculum Revisions

Revision of the curriculum is likely to be accomplished as an ongoing process in response to teacher recommendations and direct observation by the developer. Revisions directed at clarifying materials, improving in-class instructional strategies, etc., require no statistical evidence or consensus. The reasonable judgment of one teacher or one student may be sufficient basis to modify the curriculum.

Revisions based on evidence of student learning, however, are more difficult to extract from field-test results. Item analysis may help to identify weaknesses in the curriculum. Ambiguous evidence, however, complicates the developer's decisions. For example, if some classes do well and others do not, a series of judgments must be made about the reasons for different levels of student achievement to determine whether curriculum changes are warranted. In some instances, the quality of implementation may explain differential performance and reduce the importance of the less successful class. On the other hand, the quality of implementation may be related to curriculum characteristics which need improvement. This latter type of information is likely to be obtained through observations of the curriculum in use. Some differences in performance will be explainable by initial differences in student characteristics, including general in-school achievement levels, especially where randomization of the population was not accomplished. A range of explanations, including teacher motivation and interest (which influences quality of implementation); and quality of data (instruments and collection procedures) must be considered.

Two common results pose difficult choices for the developer and sponsor:

- evidence of curriculum effectiveness for some types of students (e.g., good readers) over others
- inconclusive or negative results

In the first instance, the developer has the choice of simply indicating types of students for whom the curriculum is effective or of modifying curriculum to serve a wider range of students. The magnitude of the implied revision influences the choice; large-scale revisions may require additional funding and the sponsor must therefore concur.

Inconclusive or negative results are obviously more serious. Several decisions are possible.

- Further field testing may be warranted especially if positive subjective evidence of student learning from teachers is obtained. Instrument defects may be responsible for the absence of support for observed success.
- Further revision and dissemination of the curriculum should probably be dropped if the evidence is inconclusive or negative and explanations for the findings cannot be reasonably attributed to some artifact of the field test itself.

One possible exception arises for entirely new vocational-technical areas where strong motivation exists to fill a curriculum need even if no conclusive evidence of effectiveness is obtained.

Terminating an unsuccessful effort takes rare courage. It is difficult for a developer or sponsor to accept negative findings. Further investment to improve a curriculum is more likely than termination even if it is not always justified. At minimum, the experience of failure of a curriculum development effort should be used to improve planning of new development efforts and application of funding criteria in order to avoid the costly waste of stillborn efforts.

7.2 Dissemination Decisions and Procedures

Dissemination is a crucial phase of the curriculum development process. If a curriculum test provides sufficient evidence of effectiveness to justify dissemination, this important final phase of the development process must be planned in careful detail. If the fruits of the effort are not effectively disseminated, educational improvement, the primary purpose of curriculum development, will not be achieved.

Effective dissemination can be understood only in terms of utilization or adoption of new curricula. The sole purpose of dissemination is to promote the use of effective curricula in order to improve instruction. If a curriculum is not used, obviously it will have no impact.

It is essential to differentiate dissemination from distribution, a related but lesser activity commonly mislabelled as dissemination. The distinction

may be clarified by the different measures of success for the two activities. Distribution of curriculum materials via printing and mailing of the documents may be considered successful if the documents are delivered to addressees. Dissemination of curriculum materials may be considered successful if the curriculum materials are utilized by the recipients.

Frequently, Federal or state development projects make provision only for printing and distribution of curriculum materials. Making materials available, however, does not automatically foster their use.

A growing body of research on the dissemination process offers guidance on how to promote utilization of the results of development efforts. The types of decisions which must be made and preferred strategies derive from this body of knowledge.

The types of decisions include:

1. To whom should the curriculum be disseminated?
2. By whom and through what channels should it be disseminated?
3. What methods should be employed in promoting utilization?
4. What information about curriculum effectiveness should be reported to potential users?

The developer and sponsor must devise specific strategies in response to the questions above. Utilization is likely to be promoted if the principles discussed below are observed in planning dissemination strategies.

1. User targets: Individuals who have leverage or influence on decisions made about curriculum adoption represent priority targets for dissemination. Educators in key positions, such as State Supervisors of the relevant vocational area, are probably in a better position to promote use of curriculum materials among larger numbers of users than teachers themselves. Targeting at such individuals helps to stretch limited dissemination resources.

2. Source and channels: The party chosen and the channels used to disseminate curricula can have important influence on the source of the effort. No single answer is available but two important considerations apply. First, the responsible party should have a stake in successful dissemination. The stake may be in the form of financial incentives, program performance measures, or others as appropriate. If, for example, the program under which a curriculum development project was funded were to be evaluated in terms of adoption of effective curricula, a sponsor might have powerful motivation to undertake aggressive dissemination. If, in other instances, sponsor resources for dissemination were limited, commercial publication, with its built-in financial motivation, might represent a good choice. In Part I projects, where resources for dissemination are usually limited, commercial publication should be considered as an important alternative.

3. Methods: The single most important decision regarding dissemination strategy involves the methods for reaching potential users. No substitute for direct and personal communication with users exists. Training conferences and technical assistance efforts are particularly effective because they instruct educators in the use of the curriculum. Part I projects sometimes provide for dissemination conferences which familiarize and train the professional audience in the use of the materials. Person-to-person efforts are, of course, costly but increasing concern for improving the productivity of research and development is leading to greater expenditures for this method of dissemination.

4. Reporting curriculum effectiveness: In promoting the utilization of new curricula, the developer has an obligation to furnish the evidence of effectiveness upon which the dissemination decision was based. This enables potential users to decide whether the evidence is adequate to support an adoption decision. A number of basic principle should be observed in reporting out on curriculum effectiveness to various interested audiences. These are:

   - The limitations of the results should be clearly specified.
   - The research conditions and methods should be indicated at the level of detail suitable to each specific audience.
   - The format of the report should be tailored to the audience(s).

Since the information needs of users vary, reports on field test results will vary accordingly. For example, teachers may be uninterested in detailed statistical evidence but concerned about the characteristics of the curriculum, of the settings in which it was field tested, in the judgment of other teachers, etc. Administrators may be more concerned with costs of implementation, evidence of effectiveness, etc. Sponsors will want to know the methodological details of the curriculum test.

Practical information which most potential users will want to know includes:

- For what types of students was the curriculum effective?
- What kind of learning objectives does the curriculum achieve, as evidenced in the field test?
- What special conditions, including teacher training, if any, existed in the field-test sites?
- What recommendations did teachers, students and administrators make for implementation strategies?
- Where was it tested?

Meeting the information needs of user audiences brings the developer full circle to the point of origin of the curriculum test: the identification of users and their needs for evidence. The final step in the pilot and field test process is thereby completed.

SUMMARY

1. Field-test results should be utilized in planning curriculum revisions and in targeting the curriculum at appropriate student populations.

2. Dissemination strategies for effective curricula should be planned by the sponsor and the developer.

3. To the extent possible, person-to-person strategies in the form of training and assistance should be an integral part of the dissemination/utilization process.

4. Field-test results should be reported out fully and in a form suitable to the information needs of various potential user audiences.